Ph.D. Dissertation in Chemistry

Structural studies on the β-arrestin2 in complex with a CXCR7 C-terminal phosphopeptide to reveal the active conformation of β-arrestin2, and structural and functional studies on Pgp3 from intestinal pathogen *Campylobacter jejuni*, playing a key role in helical cell shape

β-arrestin2와 CXCR7의 C-말단 인산화 펩타이드 복합체 구조 규명을 통한 β-arrestin2의 활성 기작 연구 및 장내 병원성 세균인 *Campylobacter jejuni*의 나선형 세포 모양을 결정하는 Pgp3 단백질의 구조 및 기능 연구

August 2020

The Graduate School
Seoul National University
Department of Chemistry

Kyungjin Min

Structural studies on the β-arrestin2 in complex with a CXCR7 C-terminal phosphopeptide to reveal the active conformation of β-arrestin2, and structural and functional studies on Pgp3 from intestinal pathogen *Campylobacter jejuni*, playing a key role in helical cell shape

Advisor: Hyung Ho Lee

Submitting a Ph.D. Dissertation of Chemistry
August 2020

The Graduate School
Seoul National University
Department of Chemistry

Kyungjin Min

Confirming the Ph.D. Dissertation
written by Kyungjin Min
August 2020

Chair Byeong Moon Kim (Seal)
Vice Chair Hyung Ho Lee (Seal)
Examiner Byung Woo Han (Seal)
Examiner Hyun-Woo Rhee (Seal)
Examiner Hyoun Sook Kim (Seal)

Abstract

Structural studies on the β-arrestin2 in complex with a CXCR7 C-terminal phosphopeptide to reveal the active conformation of β-arrestin2, and structural and functional studies on Pgp3 from intestinal pathogen *Campylobacter jejuni*, playing a key role in helical cell shape

Kyungjin Min

Department of Chemistry

The Graduate School

Seoul National University

The dissertation focuses on gaining a molecular level of understanding of how ligand recognition affects the alteration of protein structure and function. The dissertation is divided into two research chapters. **Chapter 1** discusses one of the important cell-signaling regulating proteins, β-arrestin 2 (βarr2) to obtain the molecular level of understanding for the G-protein coupled receptor (GPCR) recognition. **Chapter 2** discusses one of the cell-shape determining (Csd) proteins called peptidoglycan peptidase 3 (Pgp3) from *Campylobacter jejuni.* The molecular level of studies for the peptidoglycan recognition of Pgp3 protein can perhaps provide us with an insight into the mechanism for host infection of the pathogen.

In **Chapter 1**, the structural and biophysical studies on the active conformation of βarr2 will be discussed. βarrs critically regulate signaling and trafficking of GPCRs, and there are two isoforms of βarrs: βarr1 and βarr2. GPCRs are the largest family of receptors on cell membranes and comprise an important class of drug targets. To turn off G-protein-mediated GPCR signaling, GPCR kinases phosphorylate the C-terminal tail and/or intracellular loops of GPCRs, which leads to arrestin binding. To date, the high-resolution structure of active βarr1 in complex with a phosphopeptide derived from GPCR has been

revealed, but that of βarr2 remains elusive. To understand the recognition of the phosphorylated carboxyl-terminus tail of GPCR by the βarr2 protein, we determined a crystal structure of *Rattus norvegicus* βarr2 in complex with a phosphopeptide (C7pp) derived from CXCR7, a class A GPCR. The crystal structure of C7pp-bound βarr2 reveals key differences from the previously determined active conformation of βarr1. One of the key differences is that C7pp-bound βarr2 shows a relatively small inter-domain rotation. To prove the active conformation of βarr2, we used hydrogen/deuterium exchange mass spectrometry (HDX-MS) and synthetic-antibody-based conformational sensors.

In **Chapter 2**, the structural and functional studies on the Pgp3 protein from *C. jejuni* will be discussed. The helical cell shape of *C. jejuni* is important for bacterial colonization during the infection of human intestines, which is believed to be due to the specific type of crosslinking of peptidoglycan. Via the remodeling of peptidoglycan by Csd proteins, the *C. jejuni* has sustained its adeptness at colonization and pathogenesis. To understand how Csd proteins recognize peptidoglycan, we solved eight X-ray crystal structures of Pgp3 including two complex structures bound with peptidoglycan derivative substrates. In addition, functional characterization of Pgp3 by the turnover chemistry revealed that it contains both D,D-endopeptidase and D,D-carboxypeptidase activities. Catalysis is accompanied by large conformational changes upon peptidoglycan binding, whereby a loop regulates access to the active site. Furthermore, prior hydrolysis of the cross-linked peptide stem, which stems from the saccharide backbone of the peptidoglycan on one side, is a pre-requisite for its recognition and turnover by Pgp3. These analyses reveal the noncanonical nature of the transformations at the core of the events that define the morphological shape for *C. jejuni* as an intestinal pathogen.

Keywords: β-arrestin2, CXCR7, Cell-shape determining protein, Peptidoglycan, Pathogen, X-ray crystallography

Student Number: 2015-22612

Contents

List of Figures

Chapter 1.

Chapter 2.

List of Tables

Abbreviations

DDM	n-dodecyl-β-D-maltopyranoside
E. coli	*Escherichia coli*
FSEC	fluorescence detection size exclusion chromatography
h	hours
HDX-MS	hydrogen/deuterium exchange–mass spectrometry
IPTG	isopropyl-β-D-L-thiogalactopyranoside
K	kelvin degree
MALS	multi-angle light scattering
min	minute
Ni-NTA	nickel-nitrilotriacetic acid
nm	nano meter
OD	optical density
PEI	polyethylenimine
PMSF	phenylmethanesulfonylfluoride
s	seconds
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
YFP	yellow fluorescent protein

Curriculum Vitae

PERSONAL INFORMATION

Name:	Kyungjin Min
Gender:	Female
Nationality:	Korea, Republic of
Place of Birth:	Seoul, Korea
Address:	Department of Chemistry, College of Natural Sciences Seoul National University, Seoul 08826, Korea

EDUCATION

2015-2020 **Ph.D. in Chemistry**
Department of Chemistry
Seoul National University
Seoul, South Korea

2010-2012 **B.S. in Biochemistry**
Department of Chemistry
The University of Texas at Austin
Austin, Texas, USA

RESEARCH EXPERIENCE

Many times **Pohang Light Source (PLS)**
Korea (Synchrotron X-ray data collection)

2017-2018 **Single Crystal X-ray Diffractometer (SC-XRD) Operator**
National Center for Inter-University Research Facilities, NCIRF
Seoul National University
Seoul, South Korea
(Model: XtaLAB PRO, Rigaku, Japan)

2012-2013 **Research Assistant**
College of Pharmacy
The University of Texas at Austin,
Austin, Texas, USA

AWARDS

Best Young Scientist Research Award
Federation of the Korean Societies for Biomolecular Sciences (FKSBS), Oryong Hall, Gwangju Institute of Science and Technology (GIST), 2020, Oral presentation
Young Scientist Award
Korean Chemical Society (KCS)Life Science Chemistry Summer Workshop, Gangneung, Korea Institute of Science and Technology (KIST), 2018, Oral presentation

TEACHING EXPERIENCE

4. Single Crystal X-ray Diffractometer (SC-XRD) Operator, User Education Program, National Center for Inter-University Research Facilities (NCIRF), Seoul National University, 2017-2018
3. Teaching Assistant, General Chemistry Laboratory Course, Seoul National University, 2017
2. Teaching Assistant, Biochemistry Laboratory Course, Seoul National University, 2016
1. Teaching Assistant, General Chemistry Laboratory Course, Seoul National University, 2015

PRESENTATIONS & WORKSHOPS

7. *Structural and functional characterization of a peptidoglycan reshaping peptidase in intestinal pathogens*, Federation of the Korean Societies for Biomolecular Sciences (FKSBS), Oryong Hall, Gwangju Institute of Science and Technology (GIST), South Korea, Jan. 9-11, 2020, Oral presentation
6. Cryo-electron microscopy mini-symposium and workshop in Seoul National University, Institute of molecular biology and genetics, Seoul National University, South Korea, Oct. 14-15, 2019
5. *Structural insights into a cell shape modifying enzyme*, International Conference of the Korean Society for Structural Biology, Jeju, South Korea, July 4-5, 2019, Oral & Poster presentation
4. *Understanding native state of Cpf1 protein from Francisella novicida by small-angle X-ray scattering*, European Crystallographic Meetings (ECM31), Oviedo, Spain, July 3-4, 2018, Poster presentation
3. *Structural and functional characterization of a peptidoglycan modifying enzyme*, International Conference of the Korean Society for Structural Biology, Jeju, South Korea, June 27-29, 2018, Poster presentation
2. *Structural insights on a peptidoglycan modifying enzyme*, Korean Chemical Society (KCS) Life Science Chemistry Summer Workshop, Gangneung,

Korea Institute of Science and Technology (KIST), South Korea, June 21-23, 2018, Oral presentation

1. *Overexpression, crystallization and preliminary X-ray crystallography analysis of a single RNA-guided endonuclease CRISPR-Cpf1 from Francisella novicida*, International Conference of the Korean Society for Structural Biology, Samcheok, South Korea, July 3-4, 2017, Poster presentation

PUBLICATIONS

8. **Min K***, An DR*, Yoon HJ*, Rana N*, Park JS, Kim J, Lee M, Hesek D, Ryu S, Kim BM, Mobashery S, Suh SW, and Lee HH (2020) Peptidoglycan reshaping by a noncanonical peptidase for helical cell shape in *Campylobacter jejuni*, ***Nat. Commun.*** 11(1): 458 (*co-first authors)

7. **Min K***, Yoon HJ*, Park JY, Baidya M, Dwivedi H, Maharana J, Chung KY, Shukla AK, and Lee HH (2019) Crystal structure of β-arrestin 2 in complex with CXCR7 phosphopeptide (*co-first authors)
bioRxiv doi: https://doi.org/10.1101/785527
Structure
(*in press, available online:* https://doi.org/10.1016/j.str.2020.06.002)

6. **Min K**, Yoon HJ, Jo Inseong, Ha NC, Jin KS, Kim JS, and Lee HH (2018) Structural insights into the apo-structure of Cpf1 protein from *Francisella novicida*, ***Biochem. Biophy. Res. Commun.*** 498(4): 775-781

5. **Min K***, Yoon HJ*, Matsuura A, Kim YH, and Lee HH (2018) Structural basis for recognition of L-lysine, L-ornithine, and L-2,4-diamino butyric acid by lysine cyclodeaminase, ***Mol. Cells*** 41(4): doi: 10.14348/molcells (*co-first author)

4. **Min K** and Lee HH (2016) Molecular aspects of CEP55 during cytokinesis and tumorigenesis, ***Biodesign*** 4(4): 89-97

3. Choi H, **Min K**, Mikami B, Yoon H-J, and Lee HH (2016) Structural and biochemical studies reveal a putative FtsZ recognition site on the Z-ring stabilizer ZapD, ***Mol. Cells*** 39(11): 814-820

2. Koag M, **Min K**, and Lee S (2014) Structural basis for the promutagenicity of 8-bromoguanine, ***J. Biol. Chem*.** 289: 6289-6298

1. Koag M, Cheun Y, Kou Y, Ouzon-Shubeita H, **Min K**, Monzingo AF, and Lee S (2013) Synthesis and structure of 16,22-diketocholesterol bound to oxysterol-binding protein Osh4, ***Steroids*** 79: 938-944

Chapter 1.

Crystal structure of β-arrestin2 in complex with a CXCR7 phosphopeptide

Published in *Structure*

Min K*, Yoon HJ*, Park JY, Baidya M, Dwivedi H, Maharana J, Chung KY, Shukla AK, and Lee HH (2020) Crystal structure of β-arrestin 2 in complex with CXCR7 phosphopeptide. *Structure* (*in press*) (*co-first authors) (*Available online:* https://doi.org/10.1016/j.str.2020.06.002)

1.1 Introduction

1.1.1 Background of GPCR and β-arrestin signaling

G-protein-coupled receptors (GPCRs), also known as 7-transmembrane receptors (7TMs), are the largest family of receptors expressed on cell membranes and comprise an important class of drug targets. In response to ligand binding, GPCRs, which are guanine nucleotide exchange factors, activate G proteins, which then trigger downstream signaling (Figure 1.1A). To turn off the G-protein-mediated GPCR signaling, GPCR kinases phosphorylate the C-terminal tail and/or the intracellular loops of GPCRs, which leads to arrestin binding (Figure 1.1B). Although there are over 800 GPCRs in the human genome, only four arrestin genes (arrestins 1-4) have been identified. Among the four arrestin subtypes, arrestin-1 and arrestin-4 are solely related to rhodopsin and cone opsin in the visual system, whereas arrestin-2 and arrestin-3 (βarr1 and βarr2, respectively) are ubiquitously expressed; they are responsible for interactions with and the regulation of non-visual GPCRs. The interaction of βarrs with phosphorylated receptors transits βarrs to their active state, which leads to desensitization and/or internalization of GPCRs (Figure 1.1B). It is also well established that βarrs critically contribute to a range of downstream signaling responses in many different GPCRs (DeWire et al., 2007) (Figure 1.1B). In addition, βarrs are also recognized as multifunctional and versatile adaptor proteins that bind to and regulate dozens of non-receptor proteins as well (Lefkowitz et al., 2006).

The binding of GPCRs and βarrs typically consists of two kinds of interactions, including docking of the phosphorylated (or unphosphorylated) receptor tail (i.e., the carboxyl-terminus) to the N-domain of βarrs and interaction of the receptor core (i.e., the intracellular side of the receptor transmembrane bundle) with the loops on the convex side of βarrs; although,

the actual sequence of the events remains unknown (Cahill et al., 2017; Celver et al., 2002; Gimenez et al., 2012; Gurevich and Gurevich, 2004; Kovoor et al., 1999; Shukla et al., 2014; Thomsen et al., 2016) (Figure 1.2). While the primary cellular functions of βarrs are broadly conserved across different GPCRs, there is increasing evidence for receptor-specific fine-tuning of βarr functions. Although a clear mechanism for functional diversity of βarrs remains mostly elusive, it has been proposed that different patterns of receptor phosphorylation establish distinct phospho-clusters on the receptor that fine-tune the interaction pattern and conformational signatures of βarrs, resulting in specific functions (Mayer et al., 2019; Nobles et al., 2011; Reiter and Lefkowitz, 2006; Xiao et al., 2007; Yang et al., 2015). To decode how distinct phosphorylation patterns, govern the conformations and functional outputs of βarrs, it is essential to visualize the structural details of βarrs in complex with differentially phosphorylated GPCRs or their corresponding phosphopeptides.

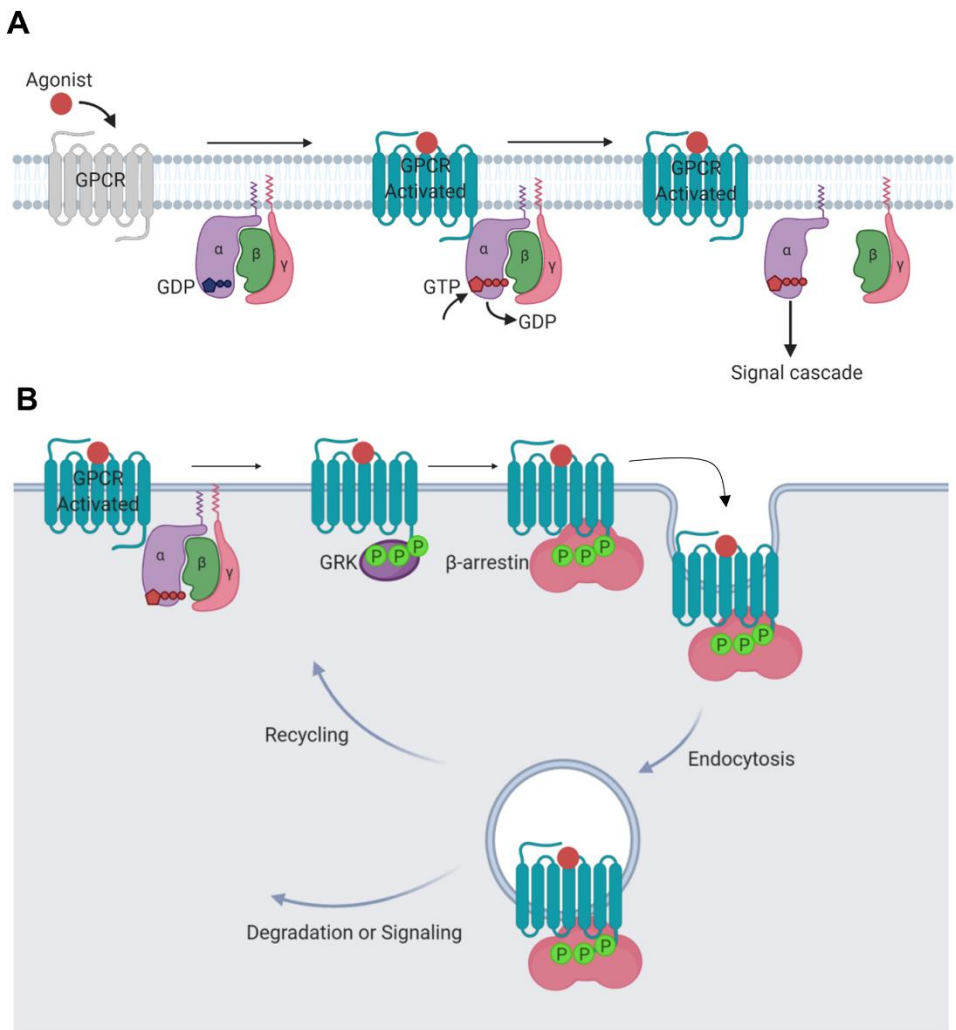


Figure 1.1 Schematic diagrams of the GPCR-activation, desensitization, and internalization pathway (modified from Liu et al., 2019; Luttrell and Lefkowitz, 2002).

(A) Schematic diagrams of the GPCR-activation pathway. In response to ligand binding, GPCRs, which are guanine nucleotide exchange factors, activate G proteins, which then trigger downstream signaling cascade. (B) The interaction of βarrs with phosphorylated GPCR by various GRKs (G protein-coupled receptor kinases) transits βarrs to their active state, which leads to desensitization and/or internalization of GPCRs.

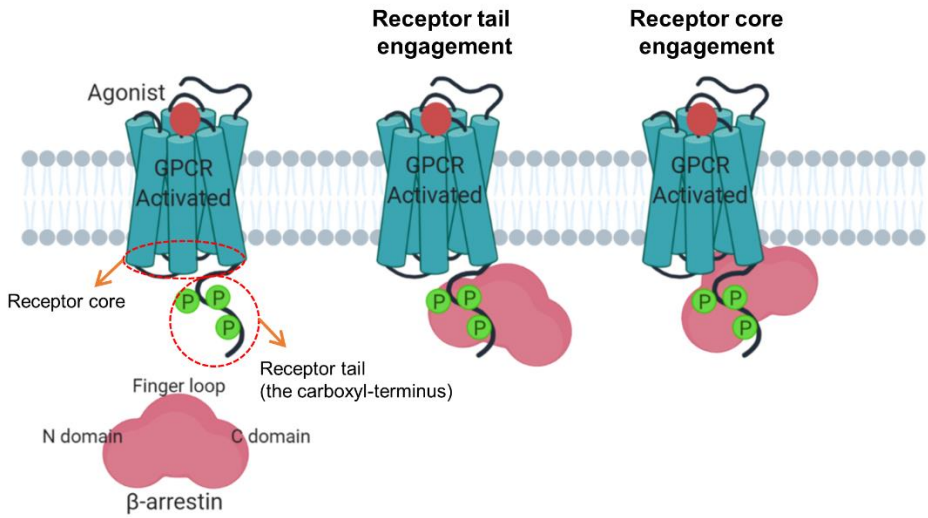


Figure 1.2 Two kinds of interactions for the binding of GPCRs and βarrs (modified from Ranjan et al., 2017).

There are two kinds of interactions for the binding of GPCRs and βarrs. One way called "receptor tail engagement" is docking of the phosphorylated (or unphosphorylated) receptor tail (i.e., the carboxyl-terminus) to the N-domain of βarrs. Another way called "receptor core engagement" is interaction of the receptor core (i.e., the intracellular side of the receptor transmembrane bundle) with the loops on the convex side of βarrs.

1.1.2 Structural studies of β-arrestin

There has been significant effort in recent years to understand the molecular mechanism of βarr activation (Table 1.1), including studies of crystal structures of pre-activated arrestin-1 (Kim et al., 2013), βarr1 in complex with the phosphorylated vasopressin receptor tail (V_2Rpp) (Shukla et al., 2013), and rhodopsin-arrestin-1 fusion protein (Kang et al., 2015; Zhou et al., 2017). These structures have revealed major conformational changes that occur upon arrestin-1 and βarr1 activation, such as significant inter-domain rotation (~ 20°), disruption of three-element (3E) and polar-core interactions, and reorientation of various loops, including the finger and lariat loops (Hirsch et al., 1999). The study of structure of V_2Rpp-bound βarr1 also confirmed a previously suggested molecular mechanism, in which the binding of the phosphorylated receptor tail to the N-domain of an arrestin displaces the carboxyl-terminus of the arrestin (Palczewski et al., 1991; Xiao et al., 2004).

Furthermore, in addition to a phosphorylation-dependent interaction, the crystal structure of rhodopsin-arrestin-1 fusion protein has also provided the structural details of a fully engaged complex, including the interface between arrestin-1 and the receptor core (Kang et al., 2015; Zhou et al., 2017). The crystal structures of inositol hexakisphosphate (IP_6)-activated βarr2, C-terminal truncated p44, and R175E mutant visual arrestin 1 also exhibited their active conformations (Cahill et al., 2017; Granzin et al., 2015; Kim et al., 2013). Single-particle negative-staining-based electron microscopy has facilitated direct visualization of the biphasic interaction between the receptor and βarr1 by capturing the partially engaged (associated through the receptor tail) and fully engaged (involving the receptor core) complexes (Shukla et al., 2014). Recently, several cryo-EM structures of βarr1 in complex with neurotensin receptor 1 or human muscarinic acetylcholine receptor showed an overall assembly that is strikingly different from that of the visual arrestin-rhodopsin

complex (Yin et al., 2019; Staus et al., 2020; Huang et al., 2020) (Table 1.1).

However, the activation of βarr2 by a phosphorylated receptor and difference between the activation of βarr2 and βarr1 or visual arr1 remain to be structurally visualized. This is particularly important considering that despite ubiquitous expression and high sequence similarity, βarr1 and βarr2 display a significant level of functional divergence (DeWire et al., 2007; Srivastava et al., 2015). For example, some GPCRs bind βarr2 with higher affinity than βarr1, while others bind both the isoforms with similar affinities (Oakley et al., 2000). Moreover, in some cases, the two isoforms of βarrs contribute differentially toward their conserved functions of receptor desensitization, endocytosis, and signaling (Srivastava et al., 2015). Additionally, for some receptors such as the bradykinin and angiotensin receptors, depletion of βarr2 results in decreased agonist-induced ERK1/2 MAP kinase phosphorylation, while depletion of βarr1 enhances the phosphorylation (Wei et al., 2003; Zimmerman et al., 2011). Thus, to fully understand βarr-mediated regulation of GPCRs and to delineate the functional divergence among βarrs, visualization of the structural details of activation of βarr2 by a phosphorylated receptor tail is essential.

Table 1.1 All arrestin structures available in Protein Data Bank.

β-arrestin	PDB ID	Identity	Coverage	Method	Gene	Construct (Status, Ligand)	Reference
1	3P2D	95%	96%	X-ray	Bos taurus β-arrestin2	Full length (Inactivated)	*J.Mol.Biol.* 2011 406 467-478
2	5TV1	95%	96%	X-ray	Bos taurus β-arrestin2	C-terminal truncated (Pre-activated, IP_6)	*Nat. Commun.* 2017 8 1427-1427
3	1JSY	75%	99%	X-ray	Bovine β-arrestin1	Full length (Inactivated)	*Biochemistry* 2002 41 3321-3328
4	6KL7	75%	97%	X-ray	Rattus norvegicus β-arrestin1	Full length (Pre-activated)	*Structure* 2020 28 314
5	3GC3	77%	96%	X-ray	Bos taurus β-arrestin1	C-terminal truncated (Inactivated, Clathrin)	*J.Biol.Chem.* 2009 284 29860-29872
6	4JQI	77%	96%	X-ray	Rattus norvegicus β-arrestin1	Full length (Activated, Nanobody&hyper-phosphorylated peptide)	*Nature* 2013 497 137-141
7	1G4M	77%	96%	X-ray	Bos taurus β-arrestin1	C-terminal truncated (Inactivated)	*Structure* 2001 9 869-880
8	6U1N	74%	96%	Cryo-EM	Rattus norvegicus β-arrestin1	C-terminal truncated (Activated, M2 human muscarinic acetylcholine receptor)	*Nature* 2020 579 297-302
9	6PWC	75%	92%	Cryo-EM	Homo sapiens β-arrestin1	C-terminal truncated (Activated, Neurotensin receptor1)	*Cell Res.* 2019 29 971-983
10	6UP7	76%	83%	Cryo-EM	Homo sapiens β-arrestin1	C-terminal truncated (Activated, Neurotensin receptor1)	*Nature* 2020 579 303-308
11	1SUJ	59%	95%	X-ray	Ambystoma tigrinum cone arrestin	Full length (Inactivated)	*J.Mol.Biol.* 2005 354 1069-1080
12	1CF1	52%	96%	X-ray	Bos taurus S-arrestin	Full length (Inactivated)	*Cell* 1999 97 257-269
13	3UGX	52%	96%	X-ray	Bos taurus S-arrestin	Full length (Inactivated)	*J.Mol.Biol.* 2012 416: 611-618
14	3UGU	56%	84%	X-ray	Bos taurus S-arrestin	C-terminal truncated (Activated)	
15	4ZRG	52%	96%	X-ray	Bos taurus S-arrestin	Full length (Inactivated)	*Sci. Rep.* 2015 5 15808-15808
16	1AYR	55%	84%	X-ray	Bos taurus S-arrestin	C-terminal truncated (Inactivated)	*Nature* 1998 391 918-921
17	4ZWJ	56%	83%	X-ray	Mouse S-arrestin	Full length (Activated, Human rhodopsin with non-phosphorylated C-terminal)	*Nature* 2015 523 561-567
18	5W0P	56%	83%	X-ray X-ray	Mouse S-arrestin	Full length (Activated, Human rhodopsin with two phosphorylated C-terminal)	*Cell* 2017 170 457-469.e13
19	6BK9	43%	86%	X-ray	Squid visual arrestin	C-terminal truncated (Inactivated)	*J.Mol.Biol.* 2018 430 4102-4118

Percentages value for identity and coverage represent sequence similarity and coverage of *rat* β-arrestin2 (residues 1-410) against other arrestin structures.

1.1.3 Background of CXCR7

Accordingly, in this study, we focused on capturing the active conformations of βarr2 in complex with phosphopeptides originating from the carboxyl-terminus of the chemokine receptor CXCR7, also referred to as atypical chemokine receptor 3 (ACKR3). CXCR7, a Class A GPCR, forms a heterodimer with another chemokine receptor, CXCR4 (Figure 1.3). The CXCR4 has a crucial function for cell migration in stem cells, leukocytes, and tumor cells (Fong et al., 2002). The expression level of CXCR4 is low or absent in heathy tissues, however, it is highly expressed in over 23 types of cancer including breast, ovarian, melanoma, and prostate cancer (Sun et al., 2010).

It has been proposed that CXCR7 acts as a "scavenger" of CXCL12, a chemokine ligand of CXCR4 (Rajagopal et al., 2010) (Figure 1.3). The CXCL12 is a chemo-attractant protein that recruits immune cells to an inflamed site of tissue (Sanchez-Martin et al., 2013). Unlike most of the GPCRs, CXCR7 only interacts with βarrs but does not display functional-coupling with heterotrimeric G-proteins (Rajagopal et al., 2010) (Figures 1.3 and 1.4). As a βarr-biased 7TM receptor, CXCR7 regulates the levels of extracellular CXCL12, the expression of CXCR4, and the CXCL12/CXCR4/G protein signaling pathway (Sanchez-Martin et al., 2013). In addition, the expression of CXCR7 was likely enhanced during the inflammation and tumor development, thus, CXCR7 is an attractive therapeutic target for autoimmune diseases and cancer (Sanchez-Martin et al., 2013). In this study, we provide an insight on the βarr-biased signaling of GPCR by capturing the active conformation of βarr2 in complex with the CXCR7 phospho-code peptide.

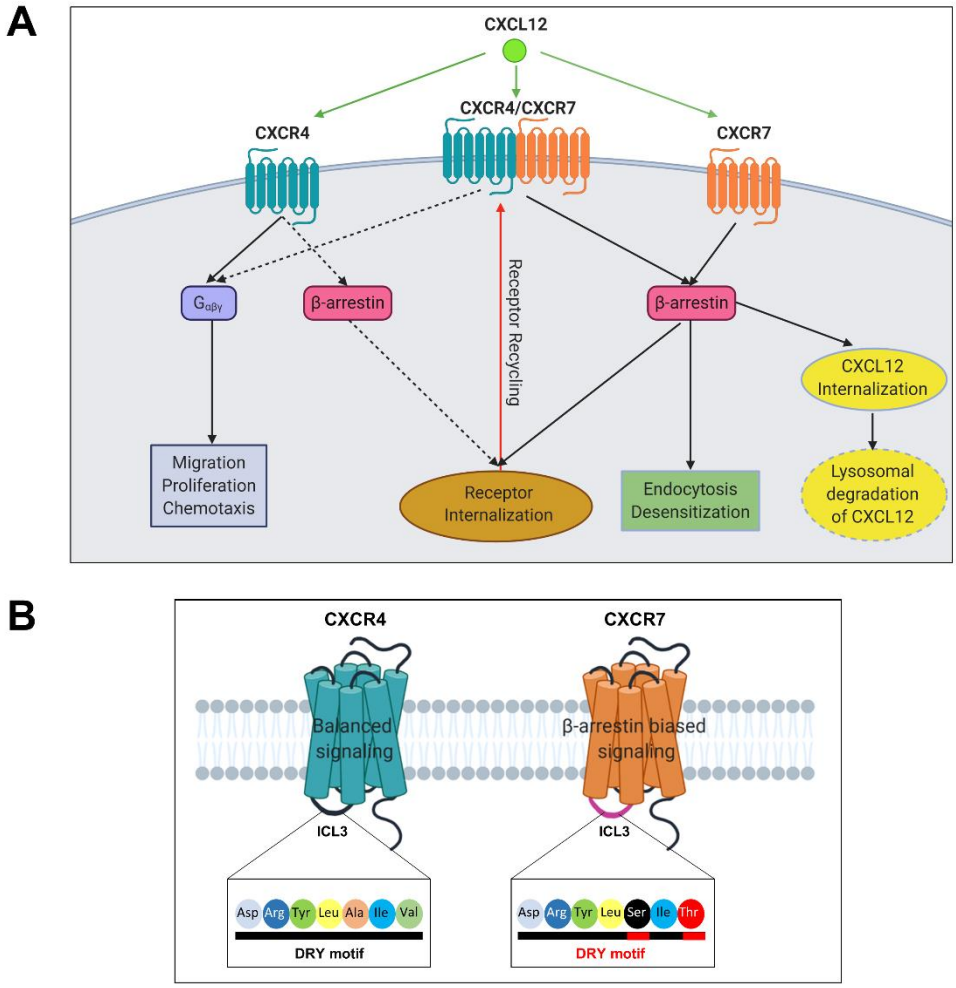


Figure 1.3 The signaling network for CXCL12 and schematic structure of CXCR4 and CXCR7 (modified from Doring et al., 2014; Asri et al., 2016).

(A) A schematic representation for the function of CXCR7. CXCL12 is a mutual ligand for CXCR4 and CXCR7. CXCR4 and CXCR7 can be formed as a heterodimer. (B) "DRYLAIV" is a common DRY motif which resides in intracellular loop 3 (ICL3) of receptor and essential for G-protein responses. The CXCR7 has modified DRY motif (DRYLSIT), thereby it recruits βarr but not G-protein.

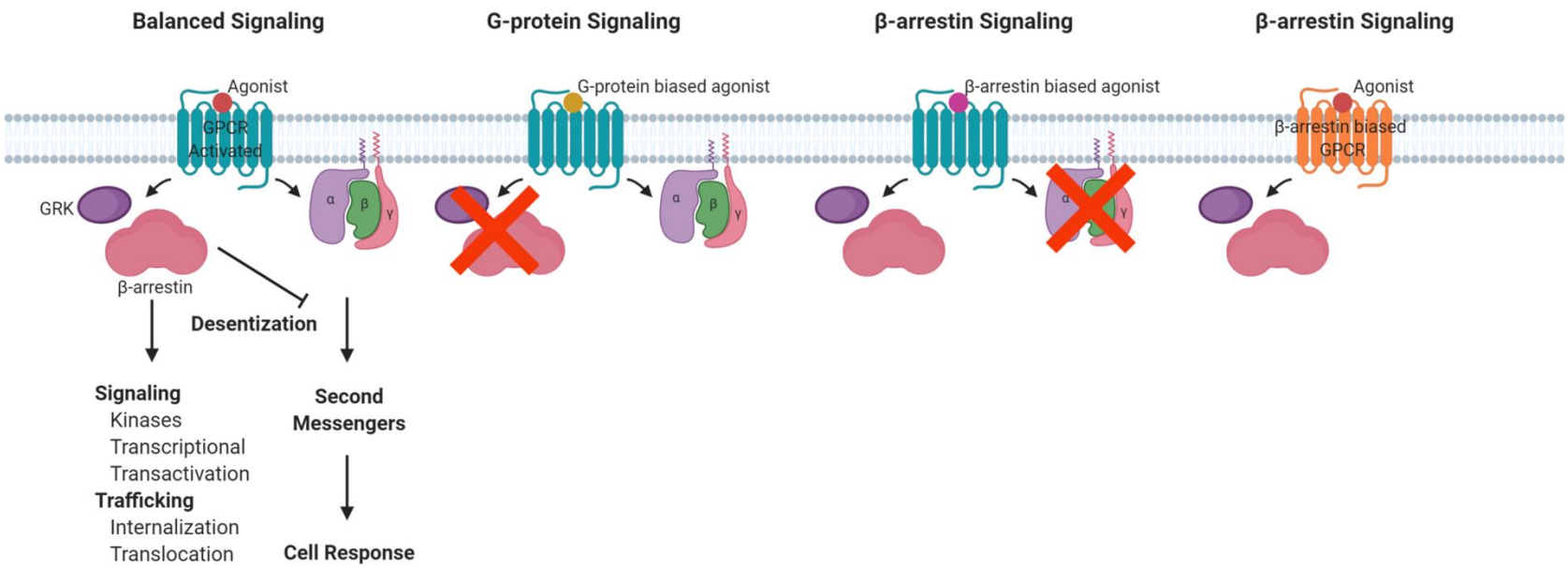


Figure 1.4 Balanced and biased signaling of GPCRs (modified from Smith and Rajagopal, 2016; Rajagopal et al., 2010).

In balanced signaling, the G-protein-mediated and βarr-mediated signaling pathways both are activated via ligand binding to the receptor. However, only one of the pathways is stimulated in G protein- or βarr-biased signaling.

1.2 Material and Methods

1.2.1 Expression and purification of recombinant *rat* β-arrestin2$_{1-356}$ and β-arrestin2$_{1-410}$

The *rat* C-terminal truncated βarr2$_{1-356}$ and wild-type βarr2$_{1-410}$ were inserted into expression vector pET28a. The plasmids were transformed into *E. coli* BL21(DE3)pLysS cells (Invitrogen), and cells harboring the plasmids were grown at 37 °C until the OD (at 600 nm) reached 0.7−1.0 in Luria Bertani (LB) broth containing 70 μg mL^{-1} chloramphenicol and 30 μg mL^{-1} kanamycin. For the crystallization experiment, cells harboring βarr2$_{1-356}$ were grown at 37 °C until the OD (at 600 nm) reached 1.0 in M9 minimal salt media (Sigma-Aldrich) containing 70 μg mL^{-1} chloramphenicol and 30 μg mL^{-1} kanamycin. Further, 0.1 mM IPTG was used to induce protein expression in the cells, after which the cells were incubated for 16 h at 16 °C. All constructs were verified by DNA sequencing.

For the isolation of the βarr2$_{1-356}$ protein fused to an N-terminal His$_6$ tag, cells were harvested by centrifugation at 5000 rpm at 4 °C for 10 min and the pellet was resuspended in ice-cold buffer A (20 mM Tris-HCl pH 8.0 and 500 mM NaCl) containing 1 mM PMSF. The cells were lysed using a microfluidizer (Microfluidics, Westwood, MA, USA) and the lysed cells were centrifuged at 15 000 rpm (Vision V506CA rotor) at 4 °C for 30 min to separate the supernatant and cell debris. The supernatant was applied to a Ni-sepharose affinity column (GE Healthcare, Little Chalfont, UK) pre-equilibrated with buffer A. Initially, the column was washed extensively with buffer A, after which the protein was eluted using buffer A containing a gradient of imidazole concentrations from 100 mM to 1 M. The eluates were desalted into buffer B (20 mM Tris-HCl pH 8.0 and 5 mM β-mercaptoethanol) containing 100 mM NaCl using a desalting column (GE Healthcare) and further purified by anion-exchange chromatography with a HiTrap Q sepharose column (GE Healthcare). The

proteins were eluted using buffer B containing 500 mM NaCl in a Q column. Further purification was performed by gel filtration on a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare), which was equilibrated with buffer B containing 200 mM NaCl. For $\beta arr2_{1–410}$, the purification steps were the same as those for the $\beta arr2_{1–356}$ construct until the application to the desalting column (GE Healthcare) (Figure 1.5). After desalting into buffer B, the protein was applied to a HiTrap heparin column (GE Healthcare) and eluted using buffer B containing 1 M NaCl. Further purification was performed by gel filtration on a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare), which was equilibrated with buffer B containing 200 mM NaCl (Figure 1.6). The homogeneity of the purified protein was assessed by polyacrylamide gel electrophoresis in the presence of 0.1% (w/v) SDS. The protein solution was concentrated to approximately 12 mg mL^{-1} using a Centricon centrifugal filter unit (Sartorius Stedim). The protein concentration was estimated by measuring the absorbance at 280 nm.

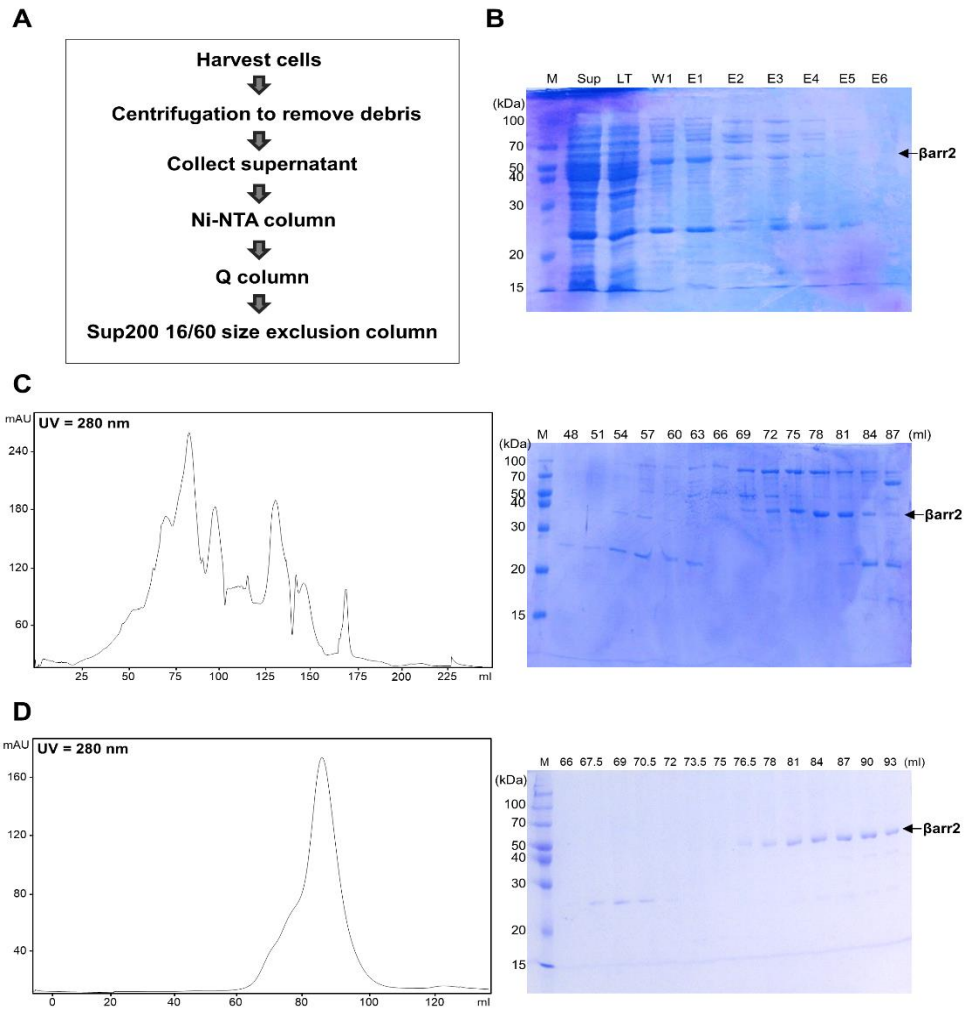


Figure 1.5 Purification of recombinant *rat* βarr2$_{1\text{-}356}$.

(A) A schematic flow chart exhibits all purification steps of recombinant *rat* βarr2$_{1\text{-}356}$. (B) The SDS-PAGE result of βarr2$_{1\text{-}356}$ after purifying by Ni-NTA affinity chromatography. Purified βarr2$_{1\text{-}356}$ was separated by 12% SDS-PAGE and stained using Coomassie brilliant blue. The protein marker is shown in lane M. The arrow indicates βarr2$_{1\text{-}356}$. (C) An FPLC elution profile and SDS-PAGE result of βarr2$_{1\text{-}356}$ purified by Q column anion-exchange chromatography. (D) An FPLC elution profile and SDS-PAGE result of βarr2$_{1\text{-}356}$ purified by Superdex 200 16/60 size exclusion chromatography.

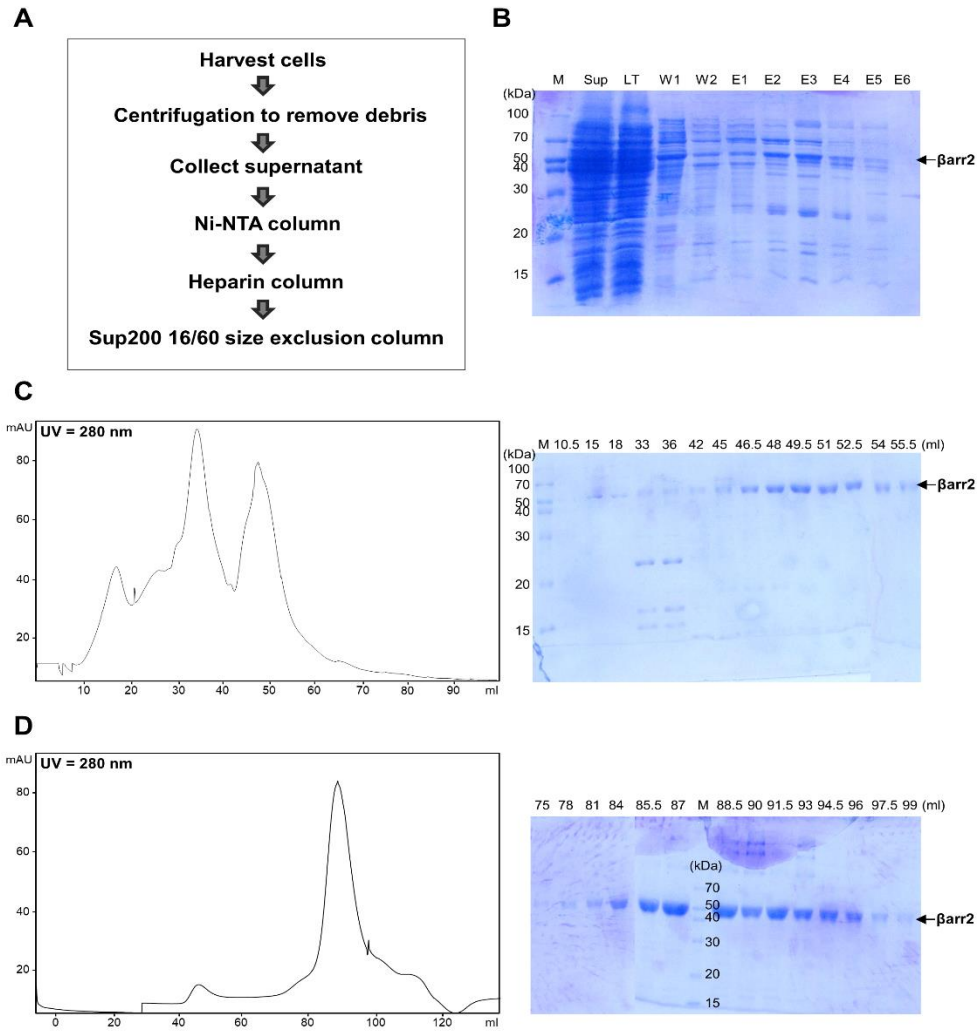


Figure 1.6 Purification of recombinant *rat* βarr2$_{1\text{-}410}$.

(A) A schematic flow chart exhibits all purification steps of recombinant *rat* βarr2$_{1\text{-}410}$. (B) The SDS-PAGE result of βarr2$_{1\text{-}410}$ after purifying by Ni-NTA affinity chromatography. Purified βarr2$_{1\text{-}410}$ was separated by 12% SDS-PAGE and stained using Coomassie brilliant blue. The protein marker is shown in lane M. The arrow indicates βarr2$_{1\text{-}410}$. (C) An FPLC elution profile and SDS-PAGE result of βarr2$_{1\text{-}410}$ purified by Heparin column chromatography. (D) An FPLC elution profile and SDS-PAGE result of βarr2$_{1\text{-}410}$ purified by Superdex 200 16/60 size exclusion chromatography.

1.2.2 Expression and purification of human $CXCR7_{1-362}$

A DNA construct containing human CXCR7 wild-type with the N-terminal FLAG epitope tag and the C-terminal YFP was generated using a pXY mammalian cell expression vector. The human CXCR7 protein was expressed in HEK293T cells (Invitrogen) by transient transfection using the mixture of branched and linear PEI (molar ratio 1:1). For transfection, 2×10^6 cells were seeded in a 100-pie plate and transfected with 8 μg CXCR7 DNA with 4 μg of the PEI mixture. The next day, the transfected cells were stimulated with 100 nM of CXCL12 (Chemotatics) for 30 min. After incubation with CXCL12, the cells were harvested and resuspended in a solubilization buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 2% (w/v) DDM (Anatrace)) supplemented with protease inhibitors (pepstatin A, PMSF, aprotinin, and leupeptin), followed by syringe homogenization. Cell membranes were solubilized for 2 h at 4 °C. The supernatant was isolated by ultracentrifugation at 100,000 × g (TLA-100 rotor, Beckman) for 1 h and incubated with anti-FLAG tag beads (FUJIFILM Wako Pure Chemical Corporation) at 4 °C overnight. The resin was washed with 100 μL of a washing buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 0.5% (w/v) DDM) and the protein was eluted with the washing buffer containing 0.5 mg mL^{-1} FLAG peptide (BIONEER) (Figure 1.7).

1.2.3 FSEC

10 μL of purified human $CXCR7_{1–362}$ with the C-terminal YFP protein at the designated concentration (1 μM) was injected into a Superose 6 Increase 5/150 GL column in 20 mM Tris pH 8.0, 200 mM NaCl, 10% glycerol, and 0.05% DDM. The fluorescence of the proteins was detected by an Agilent 1260 Infinity II fluorescence detector (Agilent Technologies), with the excitation set at 490 nm and emission at 530 nm (Figure 1.7).

1.2.4 Western blot

The CXCR7 protein samples obtained during purification steps were separated across the 12% SDS-PAGE. The separated proteins were then transferred to a polyvinylidene fluoride membranes membrane (Amersham). The protein transferred membranes were blocked overnight with the use of 2.5% (w/v) skim milk dissolving in a PBST buffer at 4 °C. The anti-FLAG (1:1000; Invitrogen) antibody was used as the primary antibody. Following that, the membranes were washed and incubated with Rabbit anti-mouse IgG antibodies (horseradish peroxidase-conjugated (HRP), 1:5000; Invitrogen) at 25 °C for 1 h. Antibody-antigen complexes detected via enhanced chemiluminescent substrate (Thermo Scientific) were captured by Amersham Imager 680 systems (GE Healthcare) (Figure 1.7).

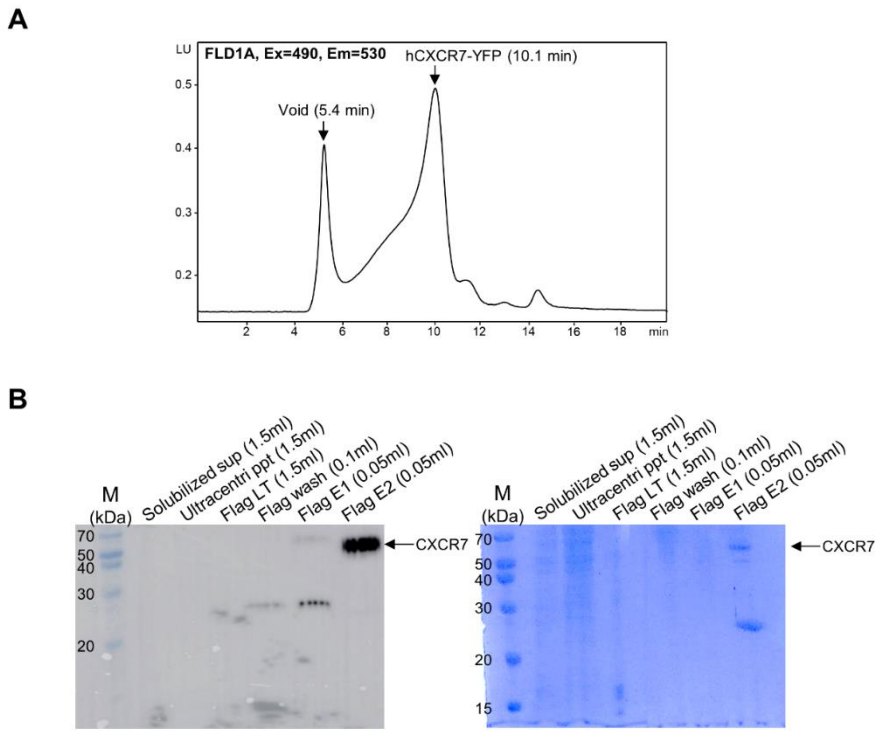


Figure 1.7 Purification result of human CXCR7$_{1\text{-}362}$.

(A) A FSEC profile is shown for solubilized human CXCR7 with the C-terminal YFP protein extracted from HEK293T cell. 'LU' stands for Luminescence Unit for fluorescence. Sample was loaded into the Superose 6 Increase 5/150 GL column. (B) Western blot (left panel) and SDS-PAGE (right panel) display solubilized and purified CXCR7 using anti-FLAG tag beads and DDM. The molecular mass marker is denoted as M (Lane 1). Lane 6 shows a purified CXCR7 band, standing at approximately 45 kDa. The sup, Ultra, and ppt stand for supernatant, Ultracentrifugation, and pellet, respectively. The LT, E1, and E2 stand for Load Through, Elution1, and Elution2, respectively.

1.2.5 Isothermal titration calorimetry (ITC)

ITC experiments were performed using Affinity ITC instruments (TA Instruments, New Castle, DE, USA) at 298 K. A 100 μM sample of $\beta arr2_{1–410}$, which was prepared in a buffer solution containing 20 mM Tris-HCl pH 8.0, and 200 mM NaCl was degassed at 295 K prior to the measurements being taken. Using a micro-syringe, 2.5 μL of 750 μM C7pp peptide solution was added at intervals of 200 s to the $\beta arr2_{1–410}$ solution in the cell with gentle stirring. A 30 μM sample of $\beta arr2_{1–410}$ was prepared in a buffer solution containing 20 mM Tris-HCl pH 8.0 and 200 mM NaCl was degassed at 295 K prior to the measurements being taken.

1.2.6 Crystallization and data collection

Before crystallization, $\beta arr2_{1–356}$ (12 mg mL^{-1}) in buffer B containing 200 mM NaCl and C7pp peptide (70 mg mL^{-1}) in 150 mM Tris pH 8.0 were mixed in a 7:1 volume ratio and incubated at 4 °C for 1 h. Crystals of the $\beta arr2_{1–356}$-C7pp complex were grown at 22 °C using sitting-drop vapor diffusion by mixing 1 μL of the protein complex solution with 1 μL of 20% (w/v) PEG 3350, 0.2 M ammonium acetate, and 0.1 M Bis-tris pH 5.5. Crystals were cryoprotected by soaking in Paratone® N oil (Sigma-Aldrich) and flash frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K in 1° oscillations at the BL26B1 beamline of a SPring-8 (Japan). Raw data were processed and scaled using the XDS program suite (Kabsch, 2010). Table 1.2 summarizes the data collection statistics. The $\beta arr2_{1–356}$-C7pp complex crystal belonged to the space group $C2_1$, with unit cell parameters of a = 91 Å, b = 127 Å, and c = 206 Å (Table 1.2).

1.2.7 Structure determination and refinement

The structure of the $\beta arr2_{1–356}$-C7pp complex was solved by the molecular

replacement method using a model of mouse visual arr1 (PDB code 5W0P). A cross-rotational search followed by a translational search was performed using the Phaser program (McCoy et al., 2007). Subsequent manual model building was performed using the COOT program (Emsley and Cowtan, 2004) and restrained refinement was performed using the REFMAC5 program (Murshudov et al., 1997). Several rounds of model building, simulated annealing, positional refinement, and individual B-factor refinement were performed. Table 1.2 lists the refinement statistics. The asymmetric unit of the $\beta arr2_{1–356}$-C7pp complex contained six molecules of $\beta arr2_{1–356}$ and peptides, where chains A, B, C, D, E, and F corresponded to $\beta arr2_{1–356}$ and chains U, V, W, X, Y, and Z corresponded to the C7pp peptide. This model included 524 water molecules and 83.3% of the residues were in the most allowed region of the Ramachandran plot. No electron density was observed for residues 175−181 in chains C and F.

Table 1.2 Data collection and refinement statistics.

Data set	βarr2 with C7pp
A. Data collection statistics	
X-ray source	SPring-8 26B
X-ray wavelength (Å)	0.97928
Space group	$C2_1$
a, b, c (Å)	91.17, 127.91, 206.04
Resolution range (Å)	50−1.95
Total / unique reflections	538,906 / 332,324
Completeness (%)	98.1 (96.3)[a]
Average I/σ (I)	70.7 (2.5)[a]
R_{merge}[b] (%)	39.0 (150.9)[a]
B. Model refinement statistics	
Resolution range (Å)	50-2.3
R_{work} / R_{free}[c] (%)	24.6 / 28.3
Number / average B-factor ($Å^2$)	
Protein nonhydrogen atoms	16,176 / 32.61
Water oxygen atoms	524 / 16.82
Peptide nonhydrogen atoms	570 / 52.28
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.004
Bond angles (°)	0.878
Protein-geometry analysis	
Ramachandran favored (%)	83.3 (1730/2077)
Ramachandran allowed (%)	14.3 (297/2077)
Ramachandran outliers (%)	2.4 (50/2077)

Footnotes for Table 1.2

[a] Values in parentheses refer to the highest resolution shell (1.95–1.98 Å).

[b] $R_{merge} = \Sigma_{hkl}\Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl}\Sigma_i I_i(hkl)_i$, where $I(hkl)$ is the intensity of reflection hkl, Σ_{hkl} is the sum over all reflections, and Σ_i is the sum over i measurements of reflection hkl.

[c] $R = \Sigma_{hkl} \mid |F_{obs}| - |F_{calc}| \mid / \Sigma_{hkl} |F_{obs}|$, where R_{free} was calculated for a randomly chosen 5% of reflections, which were not used for structure refinement, and R_{work} was calculated for the remaining reflections.

1.2.8 Hydrogen deuterium exchange mass spectrometry (HDX-MS)[1]

βarr2 protein samples were prepared in 100 μM as a final concentration in 20 mM HEPES pH 7.4 and 150 mM NaCl. For peptide binding, 500 μM of peptide was added to βarr2 and incubated for 1 h at room temperature. Hydrogen/deuterium exchange was initiated by mixing 2 μL of protein samples with 28 μL of D_2O buffer (20 mM HEPES pH 7.4, 150 mM NaCl, and 10% glycerol in D_2O) and incubating for 10, 100, 1000, or 10000 seconds on ice. At the indicated time points, the reaction was slowed down by the addition of 30 μL of ice-cold quench buffer (100 mM NaH_2PO_4 pH 2.01). For non-deuterated samples, 2 μL of protein sample was mixed with 28 μL of H_2O buffer (20 mM HEPES pH 7.4 and 150 mM NaCl in H_2O) and quenched with 30 μL of ice-cold quench buffer. The quenched samples were digested online by passing through an immobilized pepsin column (2.1 × 30 mm) at a flow rate of 100 mL/min with 0.05% formic acid in H_2O at 12 °C. Peptide fragments were subsequently collected on a C18 VanGuard trap column (1.7 mm × 30 mm) for desalting with 0.05% formic acid in H_2O. Proteins were then separated by ultra-pressure liquid chromatography over an ACQUITY UPLC C18 column (1.7 mm, 1.0 mm × 100 mm) at a flow rate of 40 mL/min with an acetonitrile gradient created by two pumps, which started with 8% B and increased to 85% B over the next 8.5 min. Mobile phase A was 0.15% formic acid in H_2O and mobile phase B was 0.15% formic acid in acetonitrile. To minimize the back exchange of deuterium to hydrogen, the sample, solvents, trap, and UPLC column were all maintained at a pH of 2.5 and 0.5 °C during analysis. Mass spectral analyses were performed with a Xevo G2 QTof equipped with a standard ESI source (Waters, Milford, MA, USA). The mass spectra were acquired in the range of m/z 100−2000 for 12 min

[1] HDX-MS was done by Ji Young Park.

in the positive ion mode. Peptides were identified in nondeuterated samples with ProteinLynx Global Server (PLGS) 2.4 (Waters, Milford, MA, USA). The following parameters were applied: monoisotopic mass, non-specific for the enzyme while allowing up to 1 missed cleavage, MS/MS ion searches, automatic fragment mass tolerance, and automatic peptide mass tolerance. Searches were performed with the variable methionine oxidation modification and the peptides were filtered with a peptide score of 6. To process the HDX-MS data, the amount of deuterium in each peptide was determined by measuring the centroid of the isotopic distribution using DynamX 2.0 (Waters, Milford, MA, USA). All measurements were performed with three independent experiments and statistical significance was analyzed by one-way ANOVA. Back-exchange levels were not corrected because the analyses compared different states.

1.3 Results

1.3.1 Generation and characterization of CXCR7 phosphopeptides

Based on a recent study that proposed the importance of different phosphorylation codes in GPCRs for βarr binding (Zhou et al., 2017), we searched a crystallizable phosphorylation-code in the carboxyl-terminus of CXCR7. To identify the phosphorylation-code, we purified CXCR7 protein from the HEK293T cell and confirmed using FSEC, SDS-PAGE, and Western blot experiments (Figure 1.7). It should be noted that previous studies have reported the functional significance of the carboxyl-terminus of CXCR7 (Hoffmann et al., 2012; Saaber et al., 2019).

1.3.2 Mapping of phosphorylation sites on CXCR7

We observed that the phosphorylation sites (Ser335, Thr338, and Thr341) in C7pp are phosphorylated upon CXCL12 stimulation by performing LC-MS analysis. We synthesized a phosphopeptide to investigate its interaction with βarr2 and any corresponding structural changes (Figure 1.8A). The peptide harbored the PxxPxxP pattern of phosphorylation, referred as to C7pp (Figures 1.8A and 1.8B), and was used to acquire an insight into the phospho-cluster-dependent structural changes of βarr2. The C7pp exhibited a binding affinity to rat βarr2, with dissociation constant (K_D) of 3.08 ± 0.3 μM as measured by isothermal titration calorimetry, displaying a monophasic binding with βarr2 (Figure 1.8C).

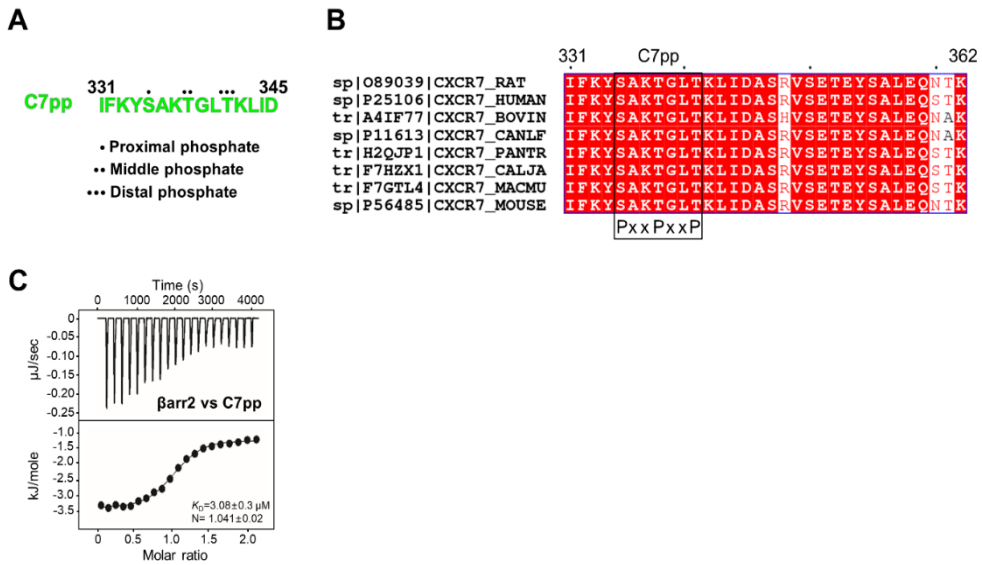


Figure 1.8 The sequence information of CXCR7 C-terminus and the CXCR7 phosphopeptide.

(A) The peptide sequence of the CXCR7 phosphopeptide referred to as C7pp (colored in green). The positions of the proximal, middle, and distal phosphates are denoted by dots. (B) Sequence alignment of CXCR7 C-terminus (residues 331-362) in several species. Multi-alignment of *R. norvegicus* CXCR7 (UniProtKB/Swiss-Prot accession number O89039) against CXCR7 from *Homo sapiens* (UniProtKB/Swiss-Prot accession number P26106), CXCR7 from *Bos taurus* (UniProtKB/Swiss-Prot accession number A4IF77), CXCR7 from *Mus musculus* (UniProtKB/Swiss-Prot accession number P56485), CXCR7 from *Canis lupus familiaris* (UniProtKB/Swiss-Prot accession number P11613), CXCR7 from *Macaca mulatta* (UniProtKB/Swiss-Prot accession number F7GTL4), CXCR7 from *Pan troglodytes* (UniProtKB/Swiss-Prot accession number H2QJP1), and CXCR7 from *Callithrix jacchus* (UniProtKB/Swiss-Prot accession number F7HZX1). The residue numbers above the sequence represent the residue numbers of rat CXCR7. Every tenth residue is marked with a black dot. Strictly (100%) conserved residues are highlighted in a red

box with characters in white and semi-conserved residues (80%) are highlighted with characters in red. A box in the sequence represents a phospho-cluster pattern (PxxPxxP). (C) The binding affinity of CXCR7 phosphopeptide with βarr2 measured with isothermal calorimetry. Purified βarr2 was incubated with increasing C7pp concentrations and the binding parameters were calculated based on the dose-response curve. The binding constant for the peptide and stoichiometry as observed in three independent experiments (N=3) is presented.

1.3.3 HDX-MS profiles of β-arrestin2 with or without co-incubation of C7pp

To understand the structural changes of βarr2 upon C7pp binding, we performed HDX-MS (Figure 1.9). HDX-MS monitors the exchange between the amide hydrogen of a protein and deuterium in the solvent, and the exchange rate is dependent on the conformational flexibility and/or solvent exposure of the amide hydrogen (Konermann et al., 2011; Skinner et al., 2012; Wales and Engen, 2006). The HDX-MS profiles of βarr2 with or without co-incubation of C7pp were analyzed, which showed that C7pp binding induced iconic changes in active arrestins. We observed increased HDX within residues 383-390 containing βXX and residues 292-301 containing the gate loop (the C-terminal part of the lariat loop), which implied release of the C-terminus and disruption of the polar core. Additionally, we observed decreased HDX within residues 119-133 containing the middle loop, residues 283-291 containing the N-terminal part of the lariat loop, and residues 305-317 containing the back loop, which implied the possible movement of the inter-domain regions (Shukla et al., 2013).

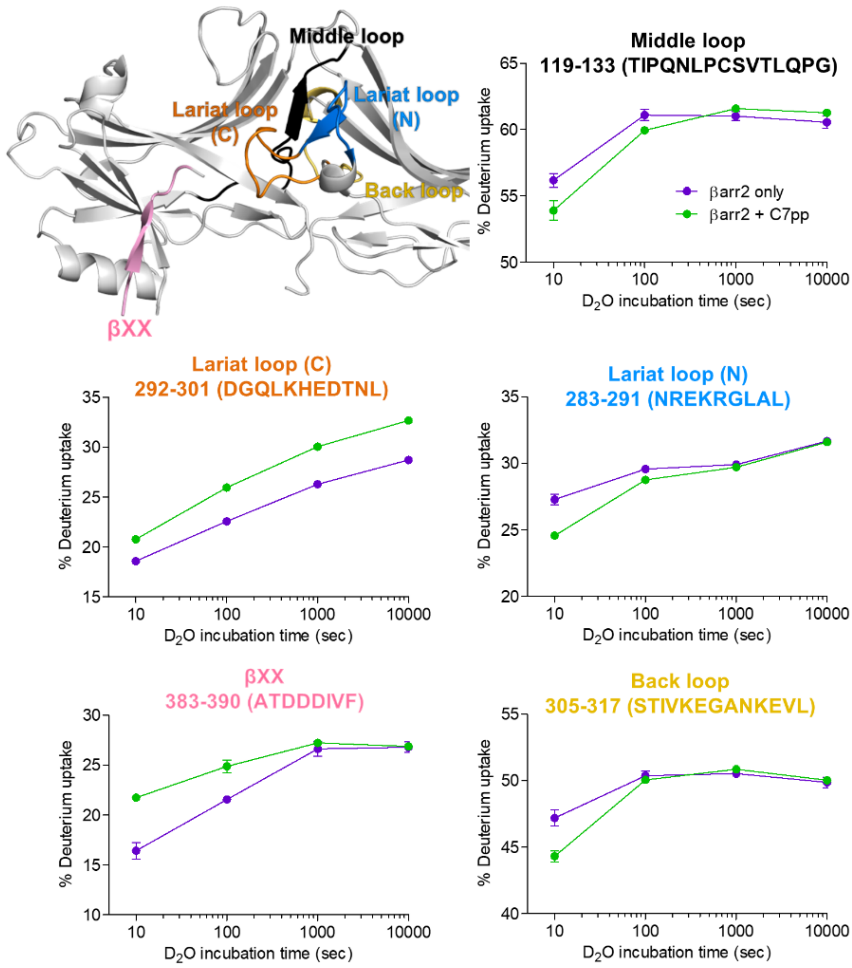


Figure 1.9 HDX-MS profile of CXCR7 phosphopeptides with βarr2.

HDX-MS profile of βarr2 upon C7pp binding. Regions with altered HDX profile are color-coded on the inactive structure of βarr2 (PDB ID: 3P2D) and the deuterium uptake plots of color-coded regions are provided. Data represent the mean ± standard error of the mean of three independent experiments. The statistical analysis was performed using one-way analysis of variance (ANOVA) followed by Tukey's post-test (*$p < 0.05$ compared to βarr2 alone). Differences smaller than 0.3 Da were not considered significant.

1.3.4 Crystal structure of *rat* β-arrestin2$_{1\text{-}356}$ in complex with CXCR7 phosphopeptide

To reveal the atomic details of C7pp-bound βarr2, we performed X-ray crystallography to obtain high-resolution structures. Although C7pp binds efficiently to full-length βarr2 (Figures 1.6, 1.8C, and 1.9), we prepared a truncated version of βarr2 that lacked the carboxyl-terminal residues 357-410 to facilitate crystallization of βarr2 in an active conformation (Figures 1.5), while all other biochemical experiments were performed using full-length βarr2. Caution was warranted while using the truncated version of βarr2, because the truncation may shift the equilibrium of βarr2 to active conformation. However, C7pp bound to the full-length βarr2 in solution (Figure 1.8C) and induced increased HDX at βXX (Figure 1.9), suggesting that C7pp can release βXX from the N-domain of βarr2. Moreover, the HDX profile of truncated βarr2 did not change upon co-incubation with C7pp (data not shown). Thus, we concluded that the C-terminal truncated version of βarr2 in complex with C7pp could represent the active conformation of βarr2 induced by C7pp binding.

We obtained a 2.3 Å refined structure of βarr2 in complex with C7pp from 1.95 Å diffraction collection data set (Figures 1.10). The crystals of C7pp-bound βarr2 appeared to be pseudo-merohedrally twinned in the $C2_1$ space group with a high R_{merge} value; thus, the structure was refined with detwinned data (Figures 1.10 and Table 1.2). The electron density map of residues 331-332 of C7pp (chain U) was not observed, while nearly all sequences of βarr2 were found to be ordered with the exception of the internal flexible regions (residues 175–181 in chain C and F, respectively) (Figure 1.11). While interpreting the structural changes in βarr2 upon C7pp binding, especially in terms of comparing them with other X-ray crystal structures of arrestins, caution was warranted during analysis of the regions involved in crystal contacts as it may sometimes lead to crystallographic artifacts. Interestingly,

the crystallographic asymmetric unit of the βarr2-C7pp complex consisted of six heterodimers of βarr2 and C7pp and revealed that the crystallographic contacts of the six molecules are not identical to each other (Figure 1.11A). Thus, we were able to find at least one solvent-exposed region amongst the six molecules for activation-dependent regions, which allowed us to confidently interpret the C7pp-induced structural changes in βarr2. Moreover, the six βarr2-C7pp molecules showed essentially similar structures overall when they were superimposed (average root-mean-square deviation of 1.14 Å for the 334 Cα atom pairs) (Figure 1.11B).

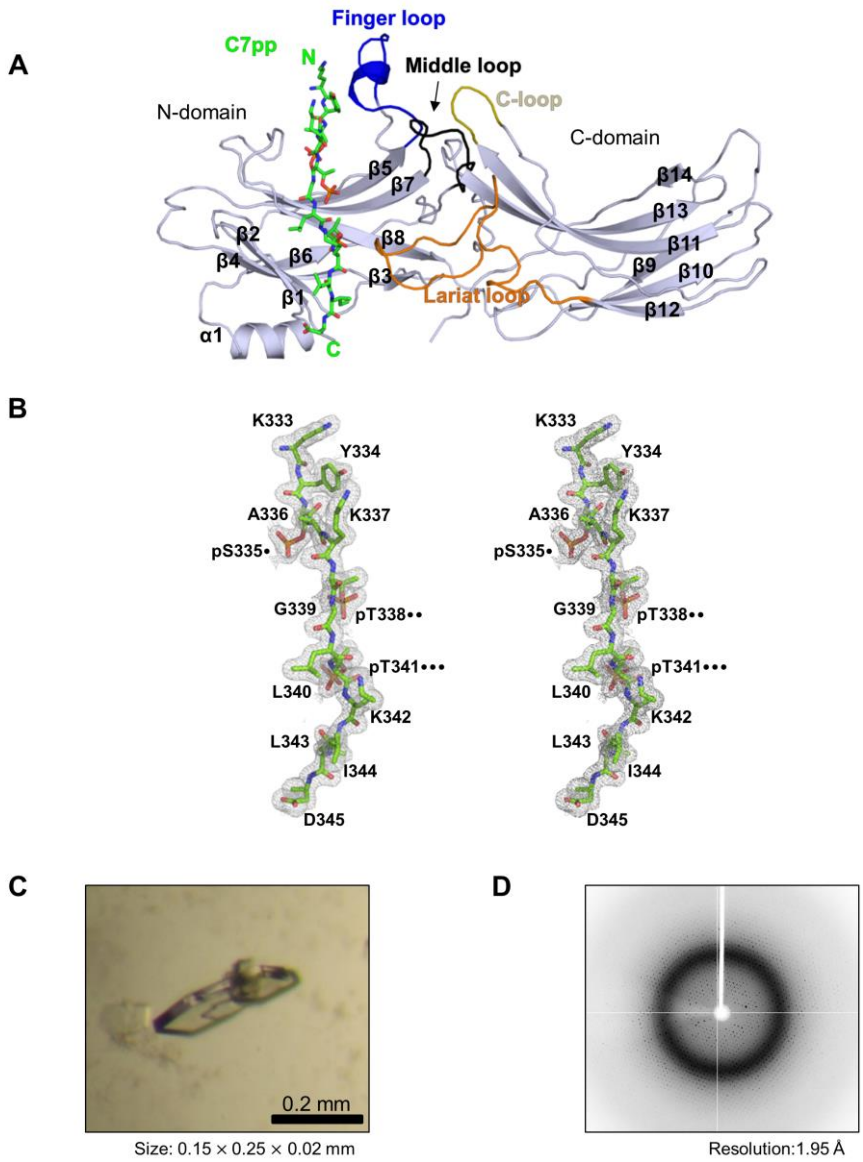


Figure 1.10 Crystal structure of βarr2 in complex with a CXCR7 phosphopeptide.

(A) Overall structural snapshot of C7pp-bound βarr2 highlighting the loop regions. The C7pp peptide is shown as green sticks and the various loops in βarr2, i.e., the finger, middle, lariat, and C loops in the central crest, are colored in blue, black, orange, and olive, respectively. (B) The stereo 2Fo - Fc map for

C7pp is drawn with a 1.0 sigma contour. The positions of the proximal, middle, and distal phosphates of the phospho- cluster (PxxPxxP) are denoted in dots. (C) A crystal picture of βarr2 in complex with CXCR7 phosphopeptide is shown. The approximate dimensions of the crystal are 0.15 mm × 0.25 mm × 0.02 mm (Width × Depth × Height). (D) The X-ray diffraction pattern for the crystal is represented.

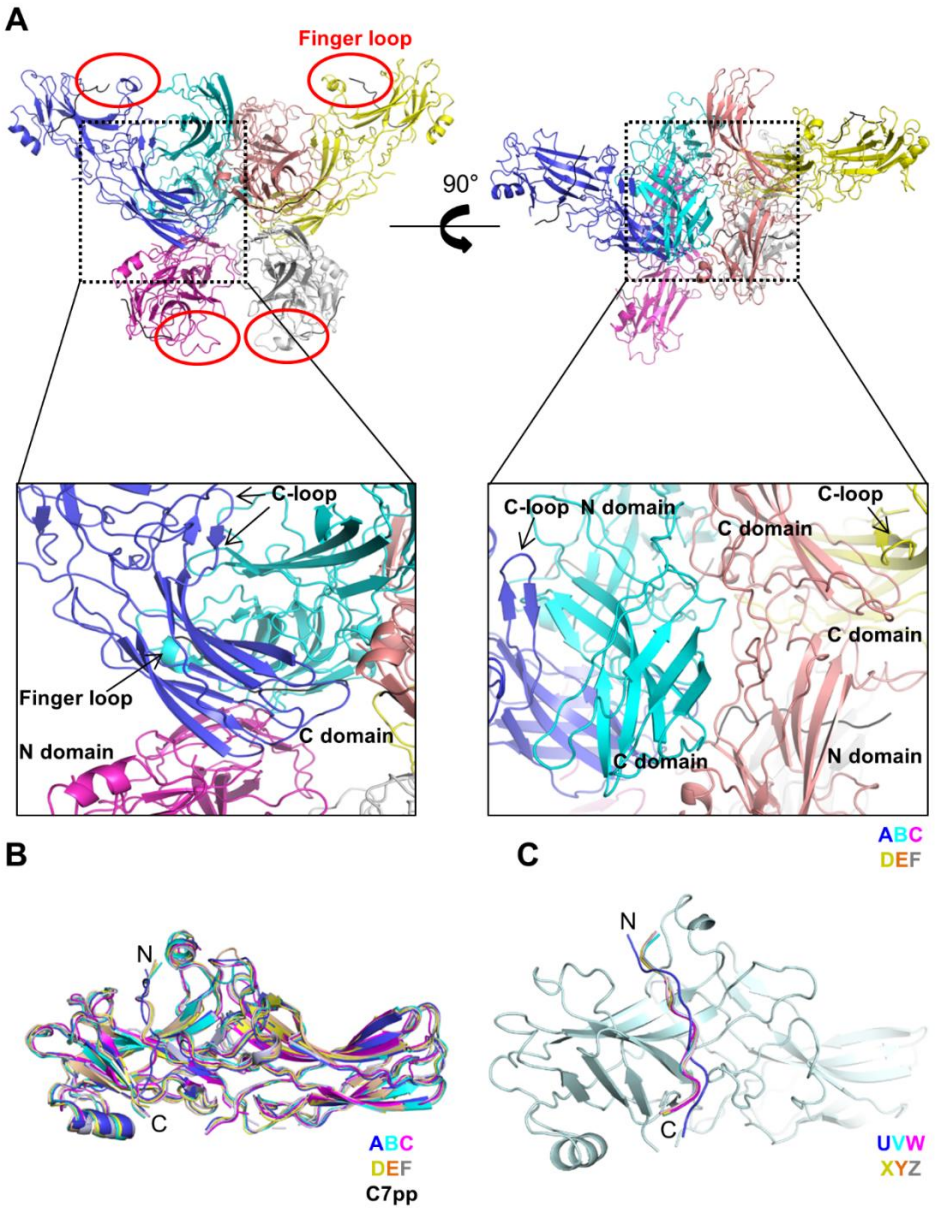


Figure 1.11 Crystal contacts of the βarr2-C7pp complex in asymmetric unit.

(A) In the $C2_1$ crystallographic lattice, six molecules of the βarr2-C7pp complex are contained in an asymmetric unit. Each chain is shown with different colors; chain A (blue), chain B (cyan), chain C (magenta), chain D (olive), chain E (orange), and chain F (gray). Dotted boxes represent the magnified crystal

contacts between chains. Red circles show the solvent exposure of the finger loop. Ribbon and surface diagrams are drawn of the six molecules rotated by 90° around the indicated axis in the right figure. The C7pp peptides are shown in black lines. (B) All six chains of the βarr2-C7pp complex in the asymmetric unit are superimposed. Each chain is shown with different colors; chain A (blue), chain B (cyan), chain C (magenta), chain D (olive), chain E (orange), and chain F (gray). (C) The N-domain of each chain is superposed to show different orientations of the C7pp peptides in each chain. Each C7pp peptide is shown as a different color; C7pp peptide U bound to chain A (blue), C7pp peptide V bound to chain B (cyan), C7pp peptide W bound to chain C (magenta), C7pp peptide X bound to chain D (olive), C7pp peptide Y bound to chain E (orange), and C7pp peptide Z bound to chain F (gray).

1.3.5 Smaller inter-domain rotation of CXCR7-β-arrestin2 compared to V_2Rpp-β-arrestin1

The structure of C7pp-bound βarr2 exhibited conformational changes similar to those in other existing active arrestin structures (βarr1, βarr2, or visual arr1), such as disruption of the 3E interaction and polar-core interaction (Figure 1.12). We used inactive βarr2 (PDB code 3P2D), IP_6-bound βarr2 (PDB code 5TV1), V_2Rpp-bound βarr1 (PDB code 4JQI), and visual arr1 R175E (PDB code 4ZRG) structures for comparison of polar core and 3E interactions with the structure of C7pp-bound βarr2. All active structures exhibited disruption of 3E and polar-core interactions (Figure 1.12). To gain further structural insights into the conformation of βarr2 induced upon its binding to C7pp, we compared our structure of C7pp-bound βarr2 to all arrestin PDB structures (Figure 1.13). We divided two groups of arrestin structures among all structures of arrestin, one group with smaller inter-domain rotation (group 1 in Figure 1.14) and the other group with larger inter-domain rotation (group 2 in Figure 1.14).

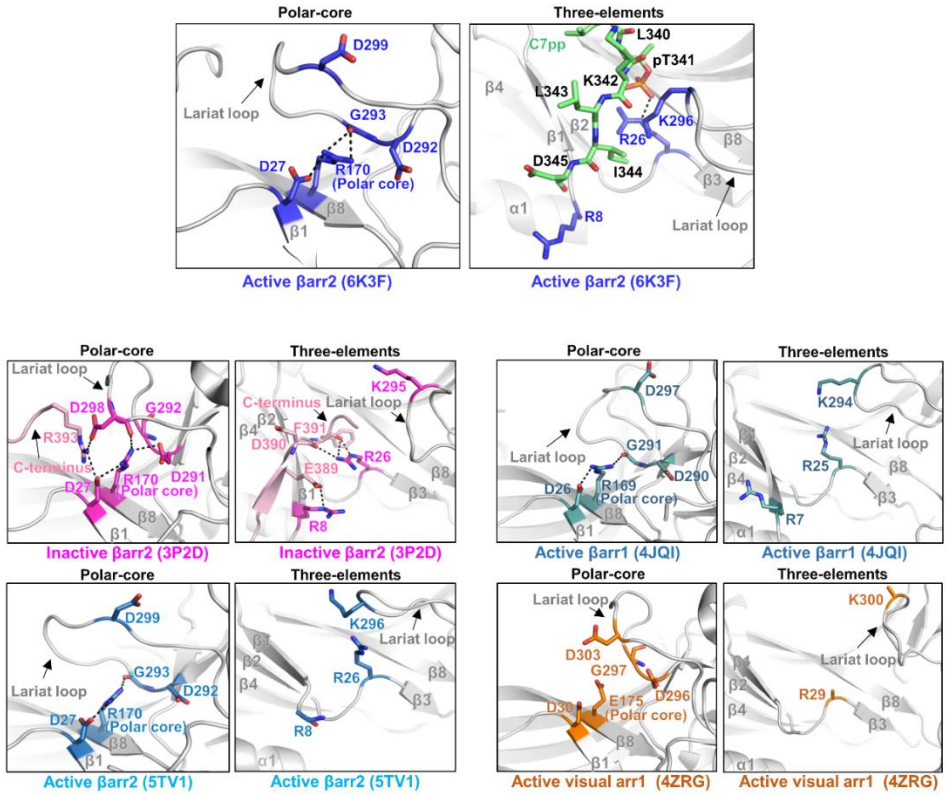
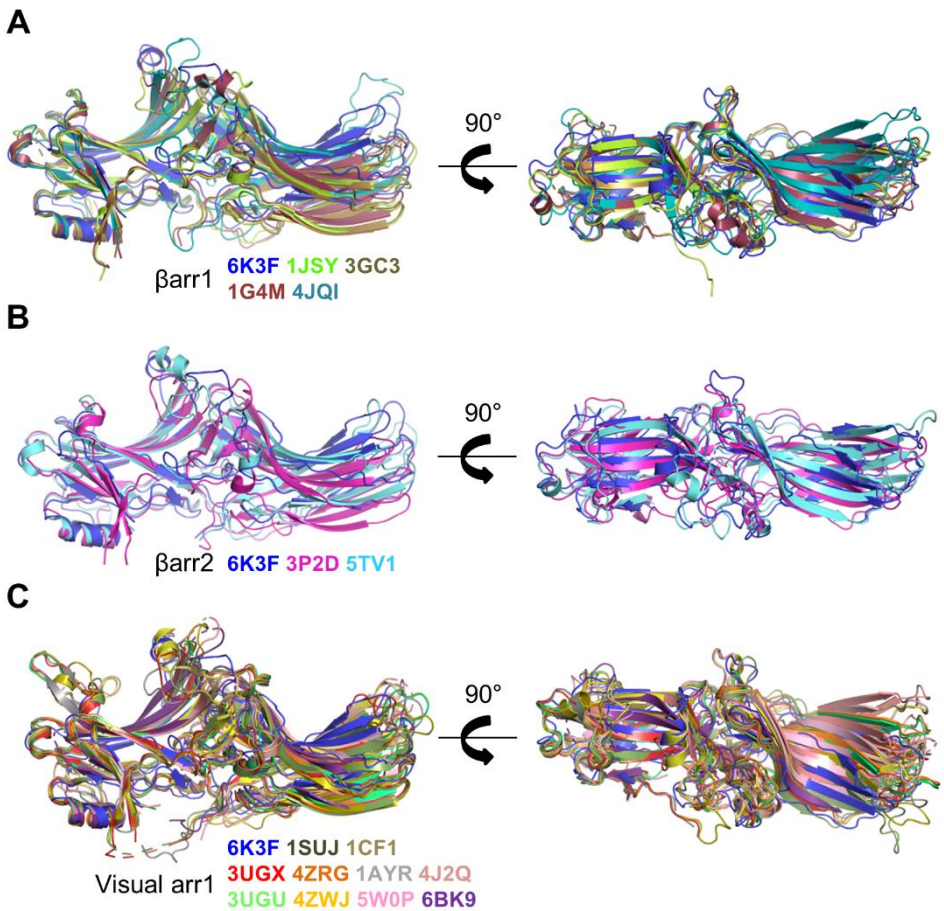


Figure 1.12 Structural comparisons of polar core and 3E interactions in arrestins.

The C7pp-bound βarr2 (6K3F, blue), Inactive βarr2 (3P2D, magenta), IP_6-bound βarr2 (5TV1, light blue), V_2Rpp-bound βarr1 (4JQI, light cyan), and R175E visual arr1 (4ZRG, orange) are shown. The C7pp is colored green and C-terminus βarr2 (3P2D, magenta) is colored pink.



D

Arrestin (species)	PDB ID	State	Inter-domain rotation angle (reference model)
βarr1 (Rat)	4JQI	Active	~20° (1G4M)
βarr2 (Bovine)	5TV1	Active	~18° (3P2D)
βarr2 (Rat)	6K3F	Active	~8° (3P2D)
Visual arr1 (Bovine)	3UGU	Active	~8° (3UGX)
Visual arr1 (Bovine)	4ZRG	Active	~8° (3UGX)
Visual arr1 (Bovine)	4J2Q	Active	~20° (1CF1)
Visual arr1 (Mouse)	4ZWJ	Active	~20° (1CF1)
Visual arr1 (Mouse)	5W0P	Active	~20° (1CF1)

Figure 1.13 Structural comparison of arrestin structures.

(A) βarr1 structures are compared and shown in different colors by superimposing the N-domains of each βarr1 structure; 6K3F (blue), 1JSY (lime), 3GC3 (dark green), 1G4M (brown), and 4JQI (light cyan) (Han et al., 2001; Kang

et al., 2009; Milano et al., 2002; Shukla et al., 2013). Superimposed ribbon diagrams of the βarr1 molecules rotated by 90° around the indicated axis in the right figure are drawn. (B) βarr2 structures are compared and shown in different colors by superimposing the N-domains of each βarr2 structure; 6K3F (blue), 3P2D (magenta), and 5TV1 (light blue). Superimposed ribbon diagrams of the βarr2 molecules rotated by 90° around the indicated axis in the right figure are drawn (Chen et al., 2017; Zhan et al., 2011). (C) Visual arr1 structures are compared and shown in different colors by superimposing the N-domains of each visual arr1 structure; 6K3F (blue), 1SUJ (dark brown), 1CF1 (olive), 3UGX (red), 4ZRG (orange), 1AYR (gray), 4J2Q (light red), 3UGU (light green), 4ZWJ (yellow), 5W0P (light pink), and 6BK9 (purple) (Bandyopadhyay et al., 2018; Granzin et al., 1998; 2012; 2015; Hirsch et al., 1999; Kang et al., 2015; Kim et al., 2013; Sutton et al., 2005; Zhou et al., 2017). Superimposed ribbon diagrams of the visual arr1 molecules rotated by 90° around the indicated axis in the right figure are drawn. (D) The inter-domain rotation angle of arrestins. Reference PDB models to calculate the inter-domain rotation angle for each arrestin structure are indicated.

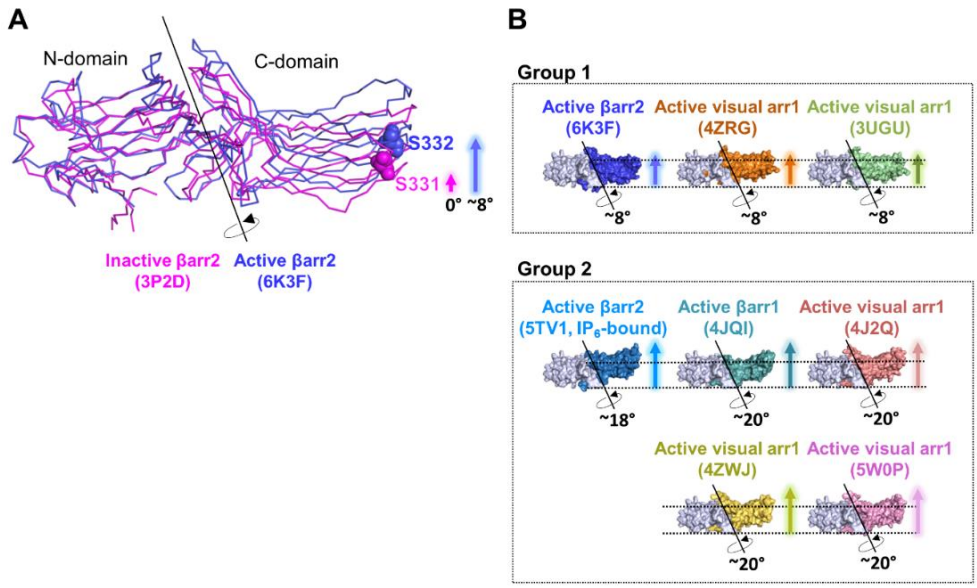


Figure 1.14 C7pp-bound βarr2 exhibits a small inter-domain rotation.

(A) The inter-domain rotation angle of βarr2 in complex with C7pp. The N-domains of active and inactive βarr2 structures are superimposed and the rotation axis is shown. The relative positions of Ser332 of the active βarr2 (PDB ID: 6K3F) are shown in ball representation as a reference for comparison. The crystal structure of βarr2 in complex with C7pp (blue) reveals an inter-domain rotation of approximately 8° compared to the inactive βarr2 structure (PDB ID: 3P2D, magenta). (B) Various inter-domain rotation angles of arrestins are shown based on previous references (4ZRG, visual arr1 R175E; 3UGU, visual arr1 p44; 5TV1, IP_6-bound βarr2; 4JQI, V_2Rpp-bound βarr1; 4J2Q, visual arr1 p44; 4ZWJ, rhodopsin-bound visual arr1; and 5W0P, rhodopsin-bound visual arr1) (Chen et al., 2017; Granzin et al., 2012; Granzin et al., 2015; Kang et al., 2015; Kim et al., 2013; Shukla et al., 2013; Zhou et al., 2017).

The inter-domain rotation angle of C7pp-bound βarr2 was found to be significantly smaller (~8°) than that of V_2Rpp-bound βarr1 (~20°) (PDB code 4JQI) (Figure 1.14). The smaller inter-domain rotation is observed in visual arr1 when Arg175 is mutated to Glu in the absence of a phosphopeptide (PDB code 4ZRG) and in C-terminal truncated visual arr1 (PDB code 3UGU) (Granzin et al., 2012; Granzin et al., 2015). Conversely, a larger inter-domain rotation (18~20°) is observed in other arrestin structures (PDB codes 4J2Q, 5TV1, 4ZWJ, and 5W0P) (Cahill et al., 2017; Kang et al., 2015; Kim et al., 2013; Zhou et al., 2017). It is interesting to note that receptor-mediated activation of βarr1 (PDB code 4JQI) and visual arr1 (PDB codes 4ZWJ and 5W0P) induces the larger inter-domain rotation, while in our study, C7pp-bound βarr2 adopted smaller inter-domain rotation. This data led us to propose two hypotheses: first, unlike the receptor-bound βarr1 of visual arr1, the receptor-bound βarr2 adopts a structure with smaller inter-domain rotation when it interacts with a phosphorylated receptor C-tail; and second, βarr2 adopts structures with various inter-domain rotations depending on the binding partners.

To test the first hypothesis, we measured the reactivity of a conformationally selective antibody fragment, Fab30, towards C7pp- and V_2Rpp-bound βarr2. Fab30 efficiently interacts with V_2Rpp-bound βarr1 and βarr2, and molecular dynamics simulations have suggested that an inter-domain rotation of more than 15° is most optimal for Fab30 reactivity (Ghosh et al., 2019). We did not observe a significant interaction of Fab30 with C7pp-bound βarr2. This was consistent with the smaller inter-domain rotation observed in the C7pp-bound crystal structure of βarr2. However, Fab30 interacted robustly with V_2Rpp-bound βarr2 (Shukla et al., 2013). These results suggested that βarr2 adopts different conformations when bound to different Rp-tails or different activation stimuli and thus led to rejection of the first

hypothesis—the smaller inter-domain rotation in C7pp-bound βarr2 structure may indicate an inherent propensity specific to βarr2 upon its activation.

An alternative hypothesis is that specific phosphorylation patterns, i.e., the number and spatial distribution of phosphates, govern the inter-domain rotation and thereby impart the corresponding functional conformation to βarr2. Although such a possibility remains to be explored further, it may explain not only the structural basis of the bar-code hypothesis but also the receptor-specific functional outcomes of βarrs. Therefore, we suggest that the current C7pp-bound βarr2 structure represents one of the active conformations that may be observed for other receptors as well, depending on the specific phosphorylation pattern. It is also tempting to suggest that this applies to βarr1 and visual arr1 as well depending on cellular and functional context. Considering that even partially engaged receptor-βarr conformations are functionally competent, for example, in terms of mediating receptor endocytosis and ERK1/2 MAP kinase activation (Cahill et al., 2017; Kumari et al., 2016), the current structure has direct implications for understanding the structural details of receptor-βarr interaction and for ensuring functional responses. It should also be noted that the structure represents the conformation of βarr2 in complex with an isolated phosphopeptide without including the receptor core interaction. It is also plausible that the core interaction may further fine-tune the conformation of βarr2, including the inter-domain rotation angle. Collectively, these data support the previously proposed model that βarr2 adopts a range of inter-domain rotations, with the domain rotation aligning the different parts of βarr2 to create a potential effector-binding site, resulting in various functional outcomes (Chen et al., 2018).

1.3.6 Distinct conformational changes of the loop regions in the C7pp-β-arrestin2 structure

The loop regions underwent conformational changes upon C7pp binding, and the structures were different in several ways from those of other active state arrestin structures (Figures 1.15A-C). First, the C7pp peptide occluded the inactive conformation of the finger loop lock, promoted outward movement, and induced a helical structure in our crystal structure (Figures 1.15A and 1.15B, left panel, blue). This was surprising because the finger loop of the βarr1-V_2Rpp complex exhibited an extended conformation (Shukla et al., 2013) (Figure 1.15C, left panel, light cyan), and the helical structure of the finger loop was often observed when the arrestin was fully docked to the GPCR core. It is also worth noting that IP_6-bound βarr2 showed helix formation in the finger loop (Figure 1.15C, left panel, light blue). Second, the middle loop structure was different and did not overlap with the structures of other arrestins (Figures 1.15A and 1.15B, middle panel). Third, the lariat loop moved most closely to the N-domain and formed van der Waal interactions with C7pp (Figures 1.15A and 1.15B, right panel). Lys296 (the corresponding residue of Lys294 in βarr1) belonging to the lariat loop moved toward C7pp, which might have provided an additional driving force for lariat loop arrangement (Figures 1.15A and 1.15B, right panel).

A similar movement was observed in the IP_6-bound βarr2 (Figure 1.15C, right panel). Based on these observations, we propose that the inter-domain rotation angle does not determine the structure of the three loops (finger, middle, and lariat loops). Given that these loops were distributed across the surface of βarr2, different phosphorylation patterns of the GPCR Rp-tail might induce distinct conformations of βarr2 in a combinatorial manner. Conversely, the C-loop, which was crucial in interacting with GPCR core, exhibited different positions depending on the inter-domain rotation angle

(Figures 1.15A and 1.15B, left panel). The arrestin structures with smaller inter-domain rotation resided in similar positions, but not in the same position (Figures 1.15A and 1.15B). Collectively, our structure does not exactly overlap with previously determined structures of arrestins, reflecting the high flexibility of arrestins.

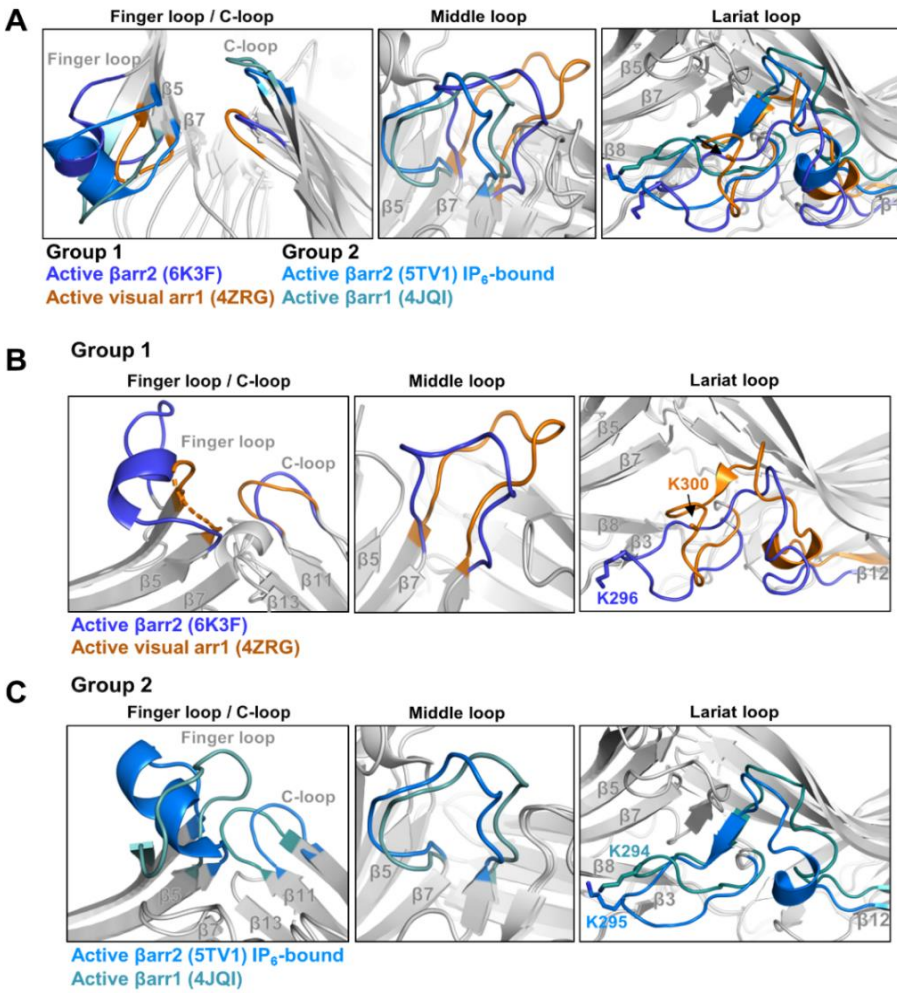


Figure 1.15 Conformational changes in various loops of βarr2 upon C7pp binding as observed in the crystal structure.

(A) Structural comparisons of the finger, middle, lariat, and C loops in C7pp-bound βarr2 (PDB code 6K3F, blue), R175E visual arr1 (PDB code 4ZRG, orange), IP_6-bound βarr2 (PDB code 5TV1, light blue), and V_2Rpp-bound βarr1 (PDB code 4JQI, light cyan). (B) Structural comparisons of the finger, middle, lariat, and C loops in C7pp-bound βarr2 (6K3F, blue) and R175E visual arr1 (4ZRG, orange). (C) Structural comparisons of the finger, middle, lariat, and C loops in IP_6-bound βarr2 (5TV1, light blue) and V_2Rpp-bound βarr1 (4JQI, light cyan).

1.3.7 Distinct binding modes of CXCR7 compared to other Rp-tails

After the examination of the conformations of six C7pp peptides in a crystallographic asymmetric unit, two types of conformations (chain U versus chains V/W/X/Y/Z) were observed with slightly different modes of βarr2 recognition (Figures 1.11C and 1.16). Therefore, there could be an ensemble of multiple conformations of C7pp when it interacts with positively charged residues distributed on the surface of βarr2 (Figure 1.16C). Given that the N-domain of βarr2 should interact with hundreds of different patterns of the GPCR Rp-tail, the complex between them might be modular, which has often been observed in disordered proteins (Miskei et al., 2017; Sente et al., 2018). The large dependence of electrostatic interactions between βarr2 and Rp-tails might allow βarr2 to pair with hundreds of GPCRs containing different phosphorylated Rp-tails.

To investigate the manner in which the binding mode of C7pp was distinct from those of V_2Rpp and the rhodopsin C-tail, we compared the conformations of these different structures (Figure 1.17). For the structural comparisons, we chose the C7pp (chain U) bound to βarr2 with chain A. It has been shown previously that the phosphopeptides overlap reasonably well when the structure of the rhodopsin-arrestin complex is superimposed with that of the βarr1-V_2Rpp complex (Zhou et al., 2017). However, when we superimposed the βarr2-C7pp complex with the βarr1-V_2Rpp complex, the overall conformations of C7pp and V_2Rpp were significantly different (Figure 1.17A). The N-terminal part of C7pp was closer to the β7/β8 loop than that of the V_2Rpp, whereas the C-terminal part of C7pp was shorter (Figure 1.17A). The N- and C-terminal parts of the V_2Rpp made a continuous β-sheet with β4 and β1, respectively, of βarr1 by anti-parallel stacking, especially in βarr1. However, those parts of C7pp did not interact directly with either β4 or β1 of βarr2 (Figure 1.17A).

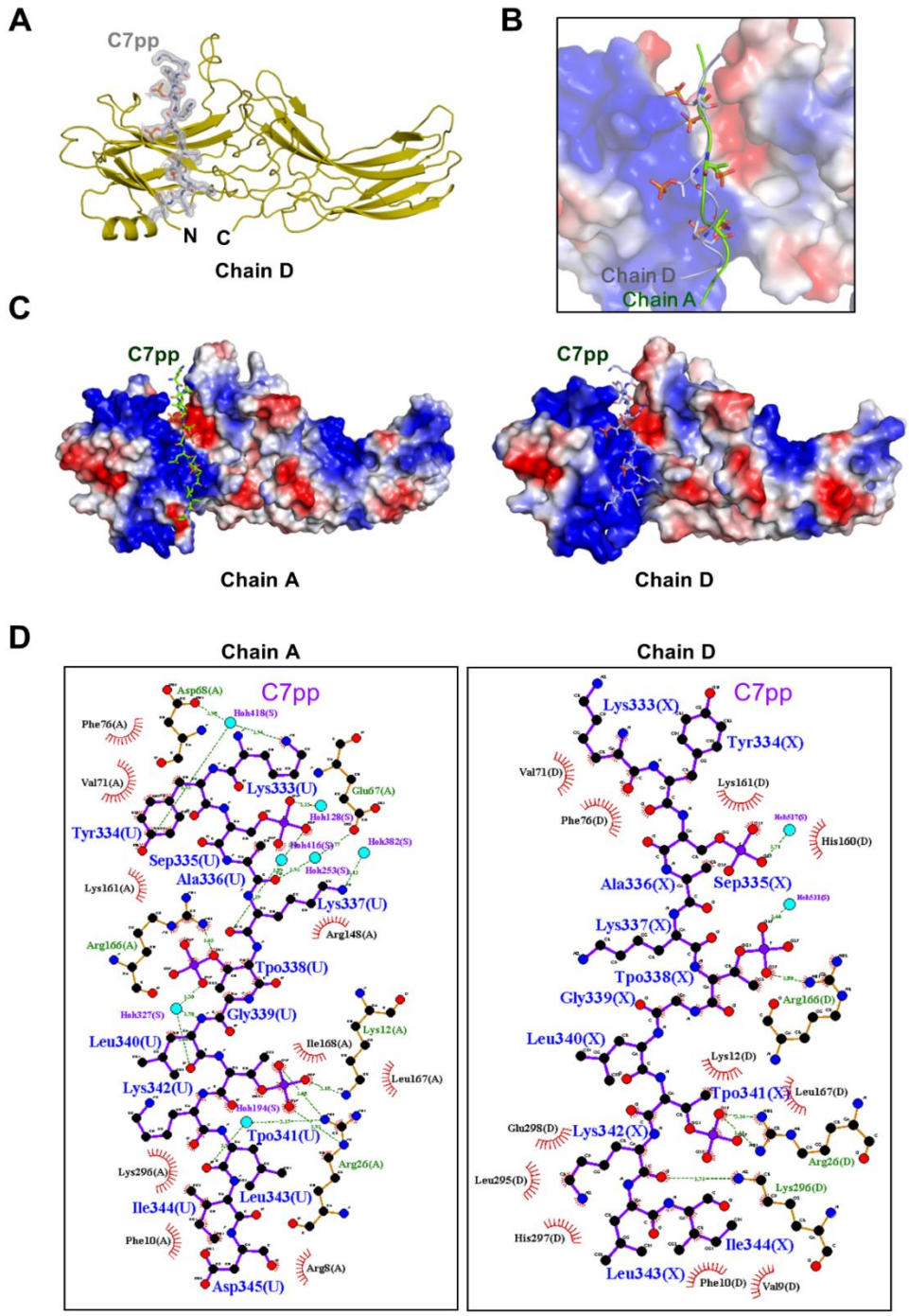


Figure 1.16 Comparison of C7pp bound to chain A and chain D.

(A) Diagram of the C7pp peptide (chain X) bound to βarr2 (chain D). The βarr2 and C7pp peptide are shown as an olive and light blue color, respectively. A 2Fo

- Fc electron density map (1.0 σ) for C7pp is shown. (B) Magnified view of both C7pp peptides (chains A and D). (C) The electrostatic potential representations for each chain (A and D) of βarr2 are shown. The C7pp peptides (chains U and X) bound to βarr2 (chains A and D) are shown as green and light blue sticks, respectively. (D) LIGPLOT Diagram. Represents the interactions between βarr2 (chains U and X) and C7pp peptides (chains A and D)(Laskowski and Swindells, 2011).

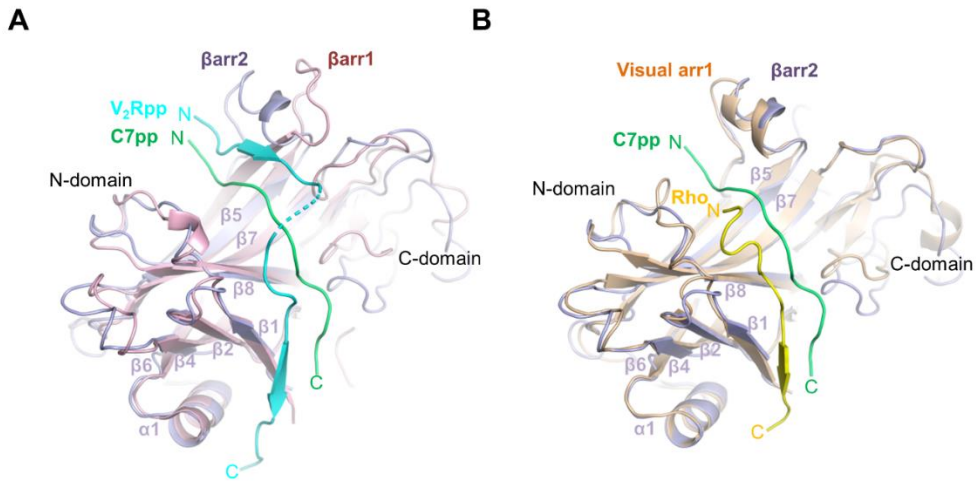


Figure 1.17 Overall distinct binding mode of C7pp with βarr2.

(A) An overall distinct binding mode of C7pp with βarr2 (6K3F, green) compared to the V_2Rpp-βarr1 complex (4JQI, light cyan). The N-domains from the crystal structures of the C7pp-βarr2 complex and V_2Rpp-βarr1 are superimposed and the respective phosphopeptides are highlighted for comparison. (B) A comparison of binding modes of C7pp with βarr2 (6K3F, green) and the rhodopsin Rp-tail with visual arr1 (5W0P, yellow) presented similar to that shown in panel A.

1.3.8 Interaction of C7pp phosphopeptide with β-arrestin2

Detailed examination of the phosphate-binding sites provided us with further insight into the different binding modes of C7pp compared to other Rp-tails. C7pp contains three phosphates, which consist of the very frequently observed phosphorylation pattern (PxxPxxP) in the GPCR C-terminus. Three positively charged pockets (pocket A, pocket B, and pocket C) might recognize the phosphorylated serine or threonine consisting of the PxxPxxP pattern (Zhou et al., 2017). pSer357 and pThr360 (the first and second phosphates) of V_2Rpp are nearly superimposable with pThr336 and pSer338 of the rhodopsin C-terminal tail, which bind to pocket A and pocket B, respectively (Zhou et al., 2017) (Figures 1.18A and 1.19). The three phosphates of C7pp make extensive contact with the positively charged residues on βarr2 (Figure 1.19). The first, second, and third phosphates (pSer335, pThr338, and pThr341) form a salt bridge with βarr2 Arg148 (2.3 Å) (box P), Arg166 (3.0 and 3.4 Å) (box A), and Arg26 (2.9 Å) (box B), respectively. Side chains of many other residues (Lys333, Lys337, Gly339, Lys342, Leu343, and Asp345), except for the phosphorylation sites (pSer335, pThr338, and pThr341), point in the opposite direction to the interface of βarr2 and C7pp (Figure 1.19); therefore, the side chains of other residues do not contribute markedly to βarr2 binding.

Instead of utilizing the same pockets (A, B, and C) in rhodopsin, the new pocket around Arg148 recognized the first phosphate (pSer335) (Figure 1.19, box P), whereas pockets A and B interacted with the second and third phosphates (pThr338 and pThr341, respectively) (Figure 1.19, boxes A and B). Therefore, the binding mode of PxxPxxP pattern was different in the βarr2-C7pp complex. We designated the newly identified pocket, which interacted with the first phosphate (pSer335), as pocket P (Figure 1.19, box P). Next, we checked whether the three pockets (P, A, and B) could accommodate the binding of the PxPxxP pattern, for which they were responsible. It appeared that

the space between the first and second phosphates could accommodate either one or two residues because the nearby Lys161 (Figure 1.19, box P), which is a strictly conserved residue (Figure 1.20), might interact with the first phosphate of the PxPxxP pattern. The phosphate sensor residues (Arg8, Lys10, Lys11, Lys107, and Lys294 in βarr1) that make contact with the V_2Rpp phosphates are not involved in the interactions with C7pp (Gurevich and Gurevich, 2004; Shukla et al., 2013) (Figures 1.18 and 1.20). The newly identified phosphate binding pocket, pocket P, might be involved in the different conformational changes of C7pp-bound βarr2 (e.g., smaller inter-domain rotation and different loop structures) compared to the V_2Rpp-bound βarr1, which requires further investigation. Together, these data suggest that various GPCR Rp-tails with different phosphorylation patterns might bind to arrestins differently, which may provide not only the strength of the interaction but also the ensuing functional outcomes.

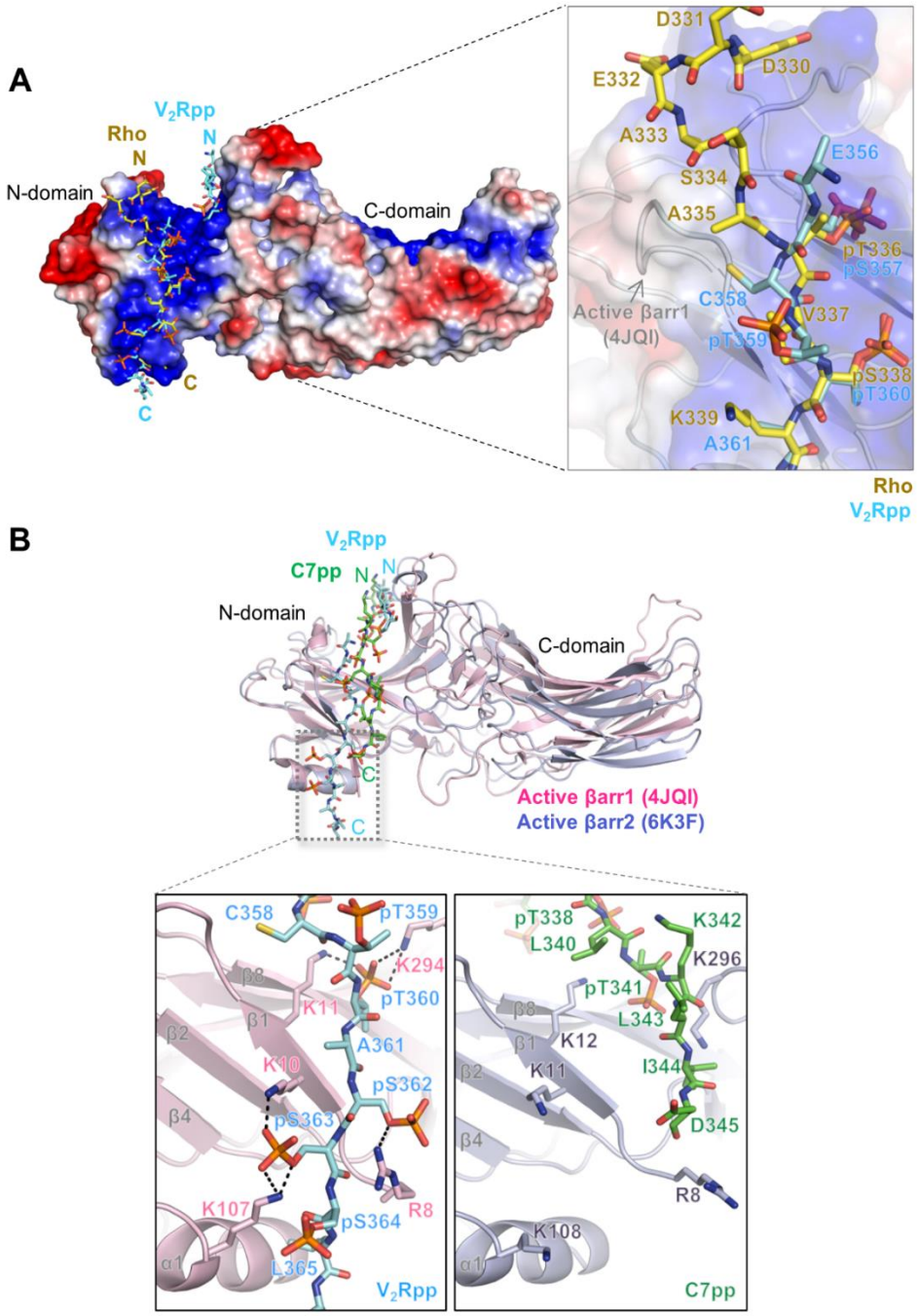


Figure 1.18 Comparisons of phosphate binding pockets in arrestins.

(A) Structural comparisons of phosphate binding sites of V_2Rpp and rhodopsin Rp-tail. The V_2Rpp (light cyan) and rhodopsin Rp-tail (yellow) were superimposed and incorporated into the electrostatic potential surface of βarr1

(4JQI). (B) Structural comparisons of phosphate binding sites of C7pp and V_2Rpp. The overall structure of the βarr2-C7pp complex (6K3F) was represented and V_2Rpp was incorporated. βarr1, βarr2, V_2Rpp, and C7pp are shown in light pink, light blue, light cyan, and green, respectively.

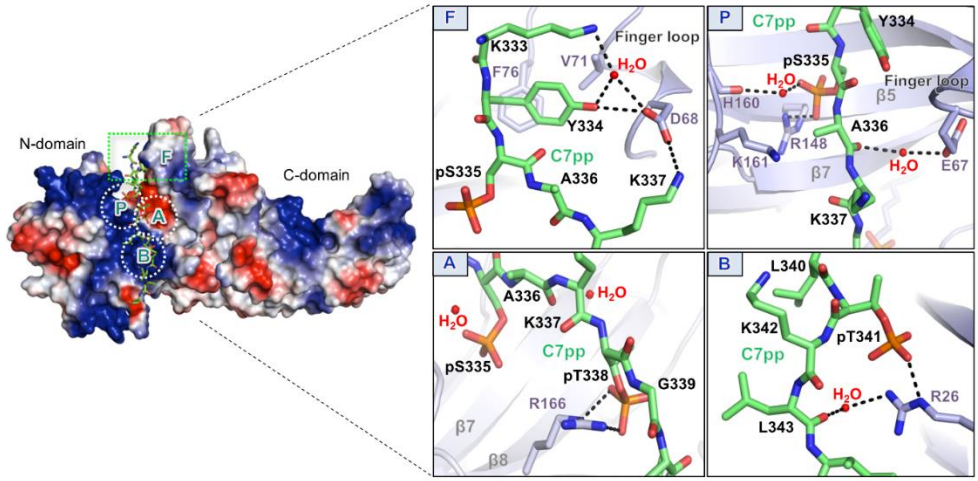


Figure 1.19 Overall binding mode of C7pp to βarr2 with specific interactions of the phosphate groups.

Surface representation of the overall electrostatic potential of the C7pp-bound βarr2 structure. C7pp is shown as green sticks. In the positive electrostatic surface of the N-domain, the four hotspots for C7pp binding are shown in the dotted rectangle or circles (F, P, A, and B). The panels on the right represent the detailed interactions at the βarr2-C7pp interface and specific interactions of the phosphates with various residues in βarr2.

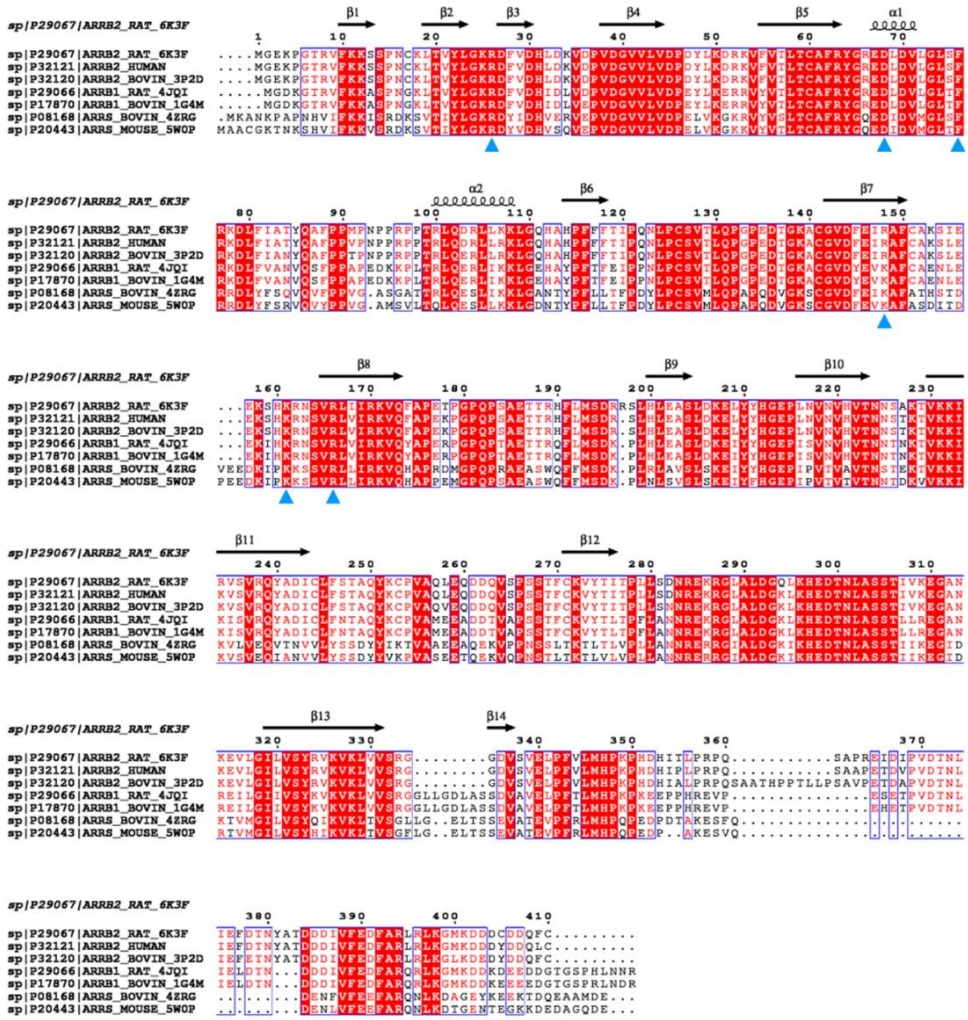


Figure 1.20 Sequence alignment of βarr2 with other arrestins.

Multi-alignment of *R. norvegicus* βarr2 (UniProtKB/Swiss-Prot accession number P29067, PDB 6K3F) against βarr2 from *H. sapiens* (UniProtKB/Swiss-Prot accession number P32121), βarr2 from *B. taurus* (UniProtKB/Swiss-Prot accession number P32120, PDB 3P2D), βarr1 from *R. norvegicus* (UniProtKB/Swiss-Prot accession number P29066, PDB 4JQI), βarr1 from *B. taurus* (UniProtKB/Swiss-Prot accession number P17870, PDB 1G4M), S-arrestin from *B. taurus* (UniProtKB/Swiss-Prot accession number P08168), and S-arrestin from *M. musculus* (UniProtKB/Swiss-Prot accession number P20443, PDB 5W0P). Secondary structural elements were assigned using PyMOL and

every tenth residue is marked with a black dot. Strictly (100%) conserved residues are highlighted in a red box with characters in white and semi-conserved residues (80%) are highlighted with characters in red. Spiral shape and arrows above the sequences denote α-helices and β-strands, respectively. The blue colored triangles represent R. norvegicus βarr2 (6K3F) residues that interact with C7pp.

1.4 Discussions

One characteristic of ACKRs, including CXCR7, is their inability to functionally couple with G-proteins while maintaining robust interaction with βarrs. Thus, it is tempting to speculate that the conformational differences observed here for C7pp-bound βarr2, compared to V_2Rpp-bound βarr1, may reflect a general feature of ACKRs. However, this possibility must be experimentally validated in the future for other ACKRs. There also exists a significant functional divergence between the two isoforms of β-arrestins, βarr1 and βarr2. Thus, it is plausible that the conformational differences between V_2Rpp-bound βarr1 and C7pp-bound βarr2 represent the mechanistic basis of this functional divergence. For example, βarrs have a direct contribution to agonist-induced ERK activation for V_2R, but for CXCR7, ERK1/2 activation was not observed (Rajagopal et al., 2010). Thus, the C7pp-bound βarr2 structure may represent one of the active conformations that are not competent for activating ERK1/2 but do support receptor endocytosis and thus ligand scavenging. However, this hypothesis requires additional experimentation in the future, including structure determination with a phosphorylated CXCR7.

In conclusion, we presented a C7pp-bound structure of βarr2 that exhibited key structural differences with the previously determined V_2Rpp-bound βarr1. These findings shed light on the functional divergence of the two βarr isoforms and underline the conformational flexibility in βarrs, which allows them to interact with multiple receptors and mediate distinct functional outcomes. Thus, our data provides useful information to obtain a better understanding of receptor-βarr interaction and signaling.

1.5 References

Asri, A., Sabour, J., Atashi, A., and Soleimani, M. (2016). Homing in hematopoietic stem cells: focus on regulatory role of CXCR7 on SDF1a/CXCR4 axis. EXCLI J. ***15***, 134-143.

Bandyopadhyay, A., Van Eps, N., Eger, B.T., Rauscher, S., Yedidi, R.S., Moroni, T., West, G.M., Robinson, K.A., Griffin, P.R., Mitchell, J.*, et al.* (2018). A novel polar core and weakly fixed c-tail in squid arrestin provide new insight into interaction with rhodopsin. J. Mol. Biol. ***430***, 4102-4118.

Cahill, T.J., 3rd, Thomsen, A.R., Tarrasch, J.T., Plouffe, B., Nguyen, A.H., Yang, F., Huang, L.Y., Kahsai, A.W., Bassoni, D.L., Gavino, B.J.*, et al.* (2017). Distinct conformations of GPCR-beta-arrestin complexes mediate desensitization, signaling, and endocytosis. Proc. Natl. Acad. Sci. U S A ***114***, 2562-2567.

Celver, J., Vishnivetskiy, S.A., Chavkin, C., and Gurevich, V.V. (2002). Conservation of the phosphate-sensitive elements in the arrestin family of proteins. J. Biol. Chem. ***277***, 9043-9048.

Chen, Q., Iverson, T.M., and Gurevich, V.V. (2018). Structural basis of arrestin-dependent signal transduction. Trends Biochem. Sci. ***43***, 412-423.

Chen, Q., Perry, N.A., Vishnivetskiy, S.A., Berndt, S., Gilbert, N.C., Zhuo, Y., Singh, P.K., Tholen, J., Ohi, M.D., Gurevich, E.V.*, et al.* (2017). Structural basis of arrestin-3 activation and signaling. Nat. Commun. ***8***, 1427.

DeWire, S.M., Ahn, S., Lefkowitz, R.J., and Shenoy, S.K. (2007). Beta-arrestins and cell signaling. Annu. Rev. Physiol. ***69***, 483-510.

Doring, Y., Pawig, L., Weber, C., and Noels, H. (2014). The CXCL12/CXCR4 chemokine ligand/receptor axis in cardiovascular disease. Front.

Physiol. ***5***, 212.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. ***60***, 2126-2132.

Fong, A.M., Premont, R.T., Richardson, Y.R.A., Lefkowitz, D.D., and Patel, D.D. (2002). Defective lymphocyte chemotaxis in beta-arrestin2-and GRK6-deficient mice. Proc. Natl. Acad. Sci. U S A ***114***, 7478-7483.

Ghosh, E., Dwivedi, H., Baidya, M., Srivastava, A., Kumari, P., Stepniewski, T., Kim, H.R., Lee, M.-H., van Gastel, J., Chaturvedi, M.*, et al.* (2019). Conformational sensors and domain-swapping reveal structural and functional differences between β-arrestin isoforms. Cell Rep. ***28***, 3287-3299.

Gimenez, L.E., Kook, S., Vishnivetskiy, S.A., Ahmed, M.R., Gurevich, E.V., and Gurevich, V.V. (2012). Role of receptor-attached phosphates in binding of visual and non-visual arrestins to G protein-coupled receptors. J. Biol. Chem. ***287***, 9028-9040.

Granzin, J., Cousin, A., Weirauch, M., Schlesinger, R., Buldt, G., and Batra-Safferling, R. (2012). Crystal structure of p44, a constitutively active splice variant of visual arrestin. J. Mol. Biol. ***416***, 611-618.

Granzin, J., Stadler, A., Cousin, A., Schlesinger, R., and Batra-Safferling, R. (2015). Structural evidence for the role of polar core residue Arg175 in arrestin activation. Sci. Rep. ***5***, 15808.

Granzin, J., Wilden, U., Choe, H.W., Labahn, J., Krafft, B., and Buldt, G. (1998). X-ray crystal structure of arrestin from bovine rod outer segments. Nature ***391***, 918-921.

Gurevich, V.V., and Gurevich, E.V. (2004). The molecular acrobatics of arrestin activation. Trends Pharmacol. Sci. ***25***, 105-111.

Han, M., Gurevich, V.V., Vishnivetskiy, S.A., Sigler, P.B., and Schubert, C. (2001). Crystal structure of beta-arrestin at 1.9 A: possible mechanism

of receptor binding and membrane Translocation. Structure ***9***, 869-880.

Hirsch, J.A., Schubert, C., Gurevich, V.V., and Sigler, P.B. (1999). The 2.8 A crystal structure of visual arrestin: a model for arrestin's regulation. Cell ***97***, 257-269.

Hoffmann, F., Muller, W., Schutz, D., Penfold, M.E., Wong, Y.H., Schulz, S., and Stumm, R. (2012). Rapid uptake and degradation of CXCL12 depend on CXCR7 carboxyl-terminal serine/threonine residues. J. Biol. Chem. ***287***, 28362-28377.

Huang, W., Masureel, M., Qu, Q., Janetzko, J., Inoue, A., Kato, H.E., Robertson, M.J., Nguyen, K.C., Glenn, J.S., Skiniotis, G.*, et al.* (2020). Structure of the neurotensin receptor 1 in complex with β-arrestin1. Nature ***579***, 303-308.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. ***66***, 125-132.

Kang, D.S., Kern, R.C., Puthenveedu, M.A., von Zastrow, M., Williams, J.C., and Benovic, J.L. (2009). Structure of an arrestin2-clathrin complex reveals a novel clathrin binding domain that modulates receptor trafficking. J. Biol. Chem. ***284***, 29860-29872.

Kang, Y., Zhou, X.E., Gao, X., He, Y., Liu, W., Ishchenko, A., Barty, A., White, T.A., Yefanov, O., Han, G.W.*, et al.* (2015). Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. Nature ***523***, 561-567.

Kim, Y.J., Hofmann, K.P., Ernst, O.P., Scheerer, P., Choe, H.W., and Sommer, M.E. (2013). Crystal structure of pre-activated arrestin p44. Nature ***497***, 142-146.

Konermann, L., Pan, J., and Liu, Y.H. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. Chem. Soc. Rev. ***40***, 1224-1234.

Kovoor, A., Celver, J., Abdryashitov, R.I., Chavkin, C., and Gurevich, V.V.

(1999). Targeted construction of phosphorylation-independent beta-arrestin mutants with constitutive activity in cells. J. Biol. Chem. ***274***, 6831-6834.

Kumari, P., Srivastava, A., Banerjee, R., Ghosh, E., Gupta, P., Ranjan, R., Chen, X., Gupta, B., Gupta, C., Jaiman, D.*, et al.* (2016). Functional competence of a partially engaged GPCR-beta-arrestin complex. Nat. Commun. ***7***, 13416.

Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model ***51***, 2778-2786.

Lefkowitz, R.J., Rajagopal, K., and Whalen, E.J. (2006). New roles for beta-arrestins in cell signaling: not just for seven-transmembrane receptors. Mol. Cell ***24***, 643-652.

Liu, X., Xu, X., Hilger, D., Aschauer, P., Tiemann, J.K.S., Du, Y., Liu, H., Hirata, K., Sun, X., Guixa-Gonzalez, R., et al. (2019). Structural insights into the process of GPCR-G protein complex formation. Cell ***177***, 1243-1251.

Luttrell, L.M., and Lefkowitz, R.J. (2002). The role of β-arrestins in the termination and transduction of G-ptorein-coupled receptor signals. J. Cell Sci. ***115***, 455-465.

Mayer, D., Damberger, F.F., Samarasimhareddy, M., Feldmueller, M., Vuckovic, Z., Flock, T., Bauer, B., Mutt, E., Zosel, F., Allain, F.H.T.*, et al.* (2019). Distinct G protein-coupled receptor phosphorylation motifs modulate arrestin affinity and activation and global conformation. Nat. Commun. ***10***, 1261.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. ***40***, 658-674.

Milano, S.K., Pace, H.C., Kim, Y.M., Brenner, C., and Benovic, J.L. (2002).

Scaffolding functions of arrestin-2 revealed by crystal structure and mutagenesis. Biochemistry ***41***, 3321-3328.

Miskei, M., Gregus, A., Sharma, R., Duro, N., Zsolyomi, F., and Fuxreiter, M. (2017). Fuzziness enables context dependence of protein interactions. FEBS Lett. ***591***, 2682-2695.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. ***53***, 240-255.

Nobles, K.N., Xiao, K., Ahn, S., Shukla, A.K., Lam, C.M., Rajagopal, S., Strachan, R.T., Huang, T.Y., Bressler, E.A., Hara, M.R.*, et al.* (2011). Distinct phosphorylation sites on the beta(2)-adrenergic receptor establish a barcode that encodes differential functions of beta-arrestin. Sci. Signal. ***4***, ra51.

Oakley, R.H., Laporte, S.A., Holt, J.A., Caron, M.G., and Barak, L.S. (2000). Differential affinities of visual arrestin, beta arrestin1, and beta arrestin2 for G protein-coupled receptors delineate two major classes of receptors. J. Biol. Chem. ***275***, 17201-17210.

Palczewski, K., Buczylko, J., Imami, N.R., McDowell, J.H., and Hargrave, P.A. (1991). Role of the carboxyl-terminal region of arrestin in binding to phosphorylated rhodopsin. J. Biol. Chem. ***266***, 15334-15339.

Rajagopal, S., Kim, J., Ahn, S., Craig, S., Lam, C.M., Gerard, N.P., Gerard, C., and Lefkowitz, R.J. (2010). Beta-arrestin- but not G protein-mediated signaling by the "decoy" receptor CXCR7. Proc. Natl. Acad. Sci. U S A ***107***, 628-632.

Ranjan, R., Dwivedi, H., Baidya, M., Kumar, M., and Shukla, A.K. (2017). Novel structural insights into GPCR-β-arrestin interaction and signaling. Trends Cell Biol. ***27***, 851-862.

Reiter, E., and Lefkowitz, R.J. (2006). GRKs and beta-arrestins: roles in

receptor silencing, trafficking and signaling. Trends Endocrinol. Metab. ***17***, 159-165.

Saaber, F., Schutz, D., Miess, E., Abe, P., Desikan, S., Ashok Kumar, P., Balk, S., Huang, K., Beaulieu, J.M., Schulz, S.*, et al.* (2019). ACKR3 regulation of neuronal migration requires ACKR3 phosphorylation, but not beta-arrestin. Cell Rep. ***26***, 1473-1488.e1479.

Sanchez-Martin, L., Sanchez-Mateos, P., and Cabanas, C. (2013). CXCR7 impact on CXCL12 biology and disease. Trends Mol. Med. ***19***, 12-22.

Scheerer, P., and Sommer, M.E. (2017). Structural mechanism of arrestin activation. Curr. Opin. Struct. Biol. ***45***, 160-169.

Sensoy, O., Moreira, I.S., and Morra, G. (2016). Understanding the differential selectivity of arrestins toward the phosphorylation state of the receptor. ACS Chem. Neurosci. ***7***, 1212-1224.

Sente, A., Peer, R., Srivastava, A., Baidya, M., Lesk, A.M., Balaji, S., Shukla, A.K., Babu, M.M., and Flock, T. (2018). Molecular mechanism of modulating arrestin conformation by GPCR phosphorylation. Nat. Struct. Mol. Biol. ***25***, 538-545.

Shukla, A.K., Manglik, A., Kruse, A.C., Xiao, K., Reis, R.I., Tseng, W.C., Staus, D.P., Hilger, D., Uysal, S., Huang, L.Y.*, et al.* (2013). Structure of active beta-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. Nature ***497***, 137-141.

Shukla, A.K., Westfield, G.H., Xiao, K., Reis, R.I., Huang, L.Y., Tripathi-Shukla, P., Qian, J., Li, S., Blanc, A., Oleskie, A.N.*, et al.* (2014). Visualization of arrestin recruitment by a G-protein-coupled receptor. Nature ***512***, 218-222.

Skinner, J.J., Lim, W.K., Bedard, S., Black, B.E., and Englander, S.W. (2012). Protein dynamics viewed by hydrogen exchange. Protein Sci. ***21***, 996-1005.

Smith, J.S., and Rajagopal, S. (2016). The β-arrestin: multifunctional regulators of G-protein-coupled receptors. J. Biol. Chem. ***291***, 8969-8977.

Srivastava, A., Gupta, B., Gupta, C., and Shukla, A.K. (2015). Emerging functional divergence of beta-arrestin isoforms in GPCR function. Trends Endocrinol. Metab. ***26***, 628-642.

Staus, D.P., Hu, H., Robertson, M.J., Kleinhenz, A.L.W., Wingler, L.M., Capel, W.D., Latorraca, N.R., Lefkowitz, R.J., and Skiniotis, G. (2020). Structure of the M2 muscarinic receptor-β-arrestin complex in a lipid nanodisc. Nature ***579***, 297-302.

Sun, X., Cheng, G., Hao, M., Zheng, J., Zhou, X., Zhang, J., Taichman, R.S., Pienta, K.J., and Wang, J. (2010). CXCL12/CXCR4/CXCR7 chemokine axis and cancer progression. Cancer Metastasis Rev. ***29***, 709-722.

Sutton, R.B., Vishnivetskiy, S.A., Robert, J., Hanson, S.M., Raman, D., Knox, B.E., Kono, M., Navarro, J., and Gurevich, V.V. (2005). Crystal structure of cone arrestin at 2.3A: evolution of receptor specificity. J. Mol. Biol. ***354***, 1069-1080.

Thomsen, A.R.B., Plouffe, B., Cahill, T.J., 3rd, Shukla, A.K., Tarrasch, J.T., Dosey, A.M., Kahsai, A.W., Strachan, R.T., Pani, B., Mahoney, J.P.*, et al.* (2016). GPCR-G protein-beta-arrestin super-complex mediates sustained G protein signaling. Cell ***166***, 907-919.

Wales, T.E., and Engen, J.R. (2006). Hydrogen exchange mass spectrometry for the analysis of protein dynamics. Mass Spectrom. Rev. ***25***, 158-170.

Wei, H., Ahn, S., Shenoy, S.K., Karnik, S.S., Hunyady, L., Luttrell, L.M., and Lefkowitz, R.J. (2003). Independent beta-arrestin 2 and G protein-mediated pathways for angiotensin II activation of extracellular signal-regulated kinases 1 and 2. Proc. Natl. Acad. Sci. U S A ***100***, 10782-10787.

Xiao, K., McClatchy, D.B., Shukla, A.K., Zhao, Y., Chen, M., Shenoy, S.K., Yates,

J.R., 3rd, and Lefkowitz, R.J. (2007). Functional specialization of beta-arrestin interactions revealed by proteomic analysis. Proc. Natl. Acad. Sci. U S A ***104***, 12011-12016.

Xiao, K., Shenoy, S.K., Nobles, K., and Lefkowitz, R.J. (2004). Activation-dependent conformational changes in {beta}-arrestin 2. J. Biol. Chem. ***279***, 55744-55753.

Yang, F., Yu, X., Liu, C., Qu, C.X., Gong, Z., Liu, H.D., Li, F.H., Wang, H.M., He, D.F., Yi, F.*, et al.* (2015). Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and (19)F-NMR. Nat. Commun. ***6***, 8202.

Yin, W., Li, Z., Jin, M., Yin, Y.-L., de Waal, P.W., Pal, K., Yin, Y., Gao, X., He, Y., Gao, J.*, et al.* (2019). A complex structure of arrestin-2 bound to a G protein-coupled receptor. Cell Res. ***29***, 971-983.

Zhan, X., Gimenez, L.E., Gurevich, V.V., and Spiller, B.W. (2011). Crystal structure of arrestin-3 reveals the basis of the difference in receptor binding between two non-visual subtypes. J. Mol. Biol. ***406***, 467-478.

Zhou, X.E., He, Y., de Waal, P.W., Gao, X., Kang, Y., Van Eps, N., Yin, Y., Pal, K., Goswami, D., White, T.A.*, et al.* (2017). Identification of phosphorylation codes for arrestin recruitment by G protein-coupled receptors. Cell ***170***, 457-469.

Zimmerman, B., Simaan, M., Akoume, M.Y., Houri, N., Chevallier, S., Seguela, P., and Laporte, S.A. (2011). Role of ssarrestins in bradykinin B2 receptor-mediated signalling. Cell. Signal. ***23***, 648-659.

Chapter 2.

Peptidoglycan reshaping by a noncanonical peptidase for helical cell shape in *Campylobacter jejuni*

Published in *Nature Communications*

Min K*, An DR*, Yoon HJ*, Rana N*, Park JS, Kim J, Lee M, Hesek D, Ryu S, Kim BM, Mobashery S, Suh SW, and Lee HH (2020) Peptidoglycan reshaping by a noncanonical peptidase for helical cell shape in *Campylobacter jejuni*, *Nat. Commun.* 11(1): 458 (*co-first authors)

2.1 Introduction

2.1.1 Biological role of cell-shape-determining proteins

The human pathogens *C. jejuni* and *Helicobacter pylori* are well-studied members of Epsilon proteobacteria, which almost exclusively is comprised of species with a curved/helical morphology (Garrity et al., 2005). Although some rod-shaped *Campylobacter* species have been described and rod-shaped variants of *C. jejuni* have been isolated, the curved/helical shape is the standard morphology of *C. jejuni* (Garrity et al., 2005). The helical shape of *C. jejuni* is important for bacterial colonization during infection, to move through the mucus layer of the gastrointestinal tract, and for entry into host cells by a corkscrew-like motility (Hermans et al., 2011). Motility of *C. jejuni* is a critical factor for host colonization and pathogenesis, with non-motile strains being severely impaired in their ability to colonize the host intestines. Thus, inhibition of proteins responsible for the helical cell shape could be useful in interference with the bacterial lifestyle and virulence (Hermans et al., 2011; Stahl et al., 2014; Wassenaar et al., 1993). *C. jejuni* infection is considered to be the most prevalent cause of bacterial diarrheal diseases worldwide, triggering severe complications such as inflammatory bowel disease, reactive arthritis, and Guillain-Barré syndrome (Blaser and Engberg, 2008; Kirk et al., 2015).

2.1.2 Cell-shape-determining proteins from *C. jejuni*

The helical shape of *C. jejuni* is believed to be due to the specific type of crosslinking of peptidoglycan (Scheffers and Pinho, 2005; Vollmer and Bertsche, 2008). Bacterial peptidoglycan is mainly comprised of a linear polysaccharide chain consisting of repeating β-(1→4)-linked N-acetylglucosamine (GlcNAc or NAG)-N-acetylmuramic acid (MurNAc or NAM) disaccharide unit, with a pentapeptide attached to the NAM (Vollmer et al., 2008). In *C. jejuni*, the pentapeptide sequence is L-Ala^{1}-γ-D-Glu^{2}-mDAP^{3}-D-Ala^{4}-D-Ala^{5}, where mDAP

refers to meso-2,6-diaminopimelate. The neighboring peptidoglycan strands are further crosslinked exclusively by the 4→3 amide linkage between the main chain of D-Ala^4 from one strand and the side chain of mDAP^3 from another strand (Figure 2.1A) (Costa et al., 1999; Meroueh et al., 2006; Sauvage et al., 2008). There are a host of enzymes that modify the peptidoglycan (Caccamo and Brun, 2018; Yang et al., 2016). In the case of *C. jejuni*, cell-shape-determining (Csd) proteins are critical (Vollmer et al., 2008). To elucidate how the peptidoglycan is remodeled by these enzymes, their structural and functional characterizations are important.

Most *C. jejuni* strains have several LytM homologs including orthologues to H. pylori Csd gene products. Two LytM homologs, Pgp1 (peptidoglycan peptidase 1) and Pgp2 (peptidoglycan peptidase 2) from *C. jejuni*, were identified as being orthologous to Csd4 and Csd6 from *H. pylori*, respectively (Esson et al., 2017; Stahl et al., 2016). Pgp1 and Pgp2 from *C. jejuni* are also responsible for cleaving peptidoglycan stem peptides by their carboxypeptidase activities (Frirdich et al., 2012; Frirdich et al., 2014). Pgp1 is a D,L-carboxypeptidase and cleaves tripeptide stems into dipeptides, whereas Pgp2 cleaves tetrapeptide stems into tripeptides as an L,D-carboxypeptidase (Frirdich et al., 2012; Frirdich et al., 2014). Mutations in the genes of Pgp1 or Pgp2 alter the muropeptide profile leading to inability to form the helical cell shape (Frirdich et al., 2012; Frirdich et al., 2014) (Figure 2.1B). In the present report, we have characterized structurally and functionally the *A8118_01115* gene product (274 amino acids) from *C. jejuni* and assigned it to the LytM protease family. We disclose that the *A8118_01115* gene product exhibits both the D,D-endopeptidase and D,D-carboxypeptidase activities. The *A8118_01115* gene product is distinct structurally from both Pgp1 and Pgp2, which we hereby designate as peptidoglycan peptidase 3 (Pgp3).

Here, we report eight X-ray crystal structures of Pgp3, which show the

protein in two unique conformations with an open and a closed active site. The structures are: (i) tartrate-bound wild-type (open form), (ii) citrate-bound wild-type (two structures, closed form), (iii) H216A mutant (open form), (iv) H216A mutant with tartrate (closed form), (v) H247A mutant (open form), (vi) H247A mutant with pentapeptide, Ac-L-Ala^1-γ-D-Glu^2-mDAP^3-D-Ala^4-D-Ala^5 (open form) (Figure 2.1A), and (vii) H247A mutant with cross-linked muramyl tetra-tri peptide, NAM-L-Ala^1-γ-D-Glu^2-mDAP^3-D-Ala^4-mDAP^3-γ-D-Glu^2-Ala^1-NAM (open form) (Figure 2.1A). Tetra-tri peptide is a minimalist D,D-cross-linked substrate. Despite our efforts to crystallize the wild-type enzyme with peptidoglycan substrates, the crystals turned over the substrates, which made the task impossible. The structures of the complexes were determined with an inactive mutant variant. This report describes the enzymatic activity of Pgp3, its conformational states, and domain architecture. The analysis provides insight into the catalytic mechanism of this important bacterial enzyme.

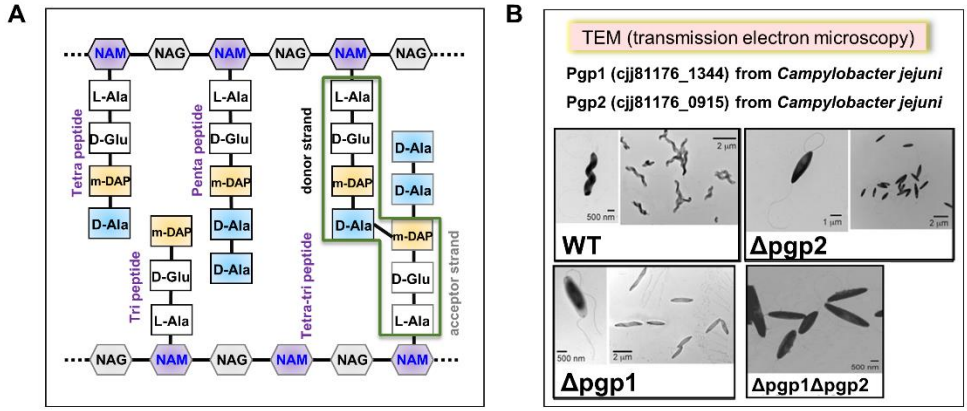


Figure 2.1 A schematic diagram for the peptidoglycan structure and functional characterizations of Pgp proteins (modified from Wyckoff et al., 2012; Frirdich et al., 2014).

(A) A schematic diagram is shown for the peptidoglycan structure from *C. jejuni*. (B) *C. jejuni* ATCC 33560 wild-type, Δpgp1, and Δpgp2 were negatively stained with uranyl acetate and then observed by using TEM. Scale bars represent 0.5 μm or 1 μm or 2 μm.

2.2 Material and Methods

2.2.1 Reaction of Pgp3 with synthetic peptidoglycans detected by liquid-chromatography mass spectrometry[2,3]

The reactions of Pgp3 with seven synthetic peptidoglycan substrates (2-8, Figure 2.2 for chemical structures) were carried out in 50 mM Tris, pH 7.5 at 23 °C and were stopped by the addition of trifluoroacetic acid. The resultant mixture was analyzed by LC/MS whose conditions were reported previously (Lee et al., 2015). Seven synthetic substrates and five product standards were synthesized by the methodology reported previously (Lee et al., 2015; Lee et al., 2010; Hesek et al., 2009; Lee et al., 2017).

2.2.2 Reaction of Pgp3 with sacculus[2]

The sacculus of *E. coli* K12 was isolated by the general procedure described previously (Suvorov et al., 2008). The reaction of Pgp3 with sacculus was carried out in 50 mM Tris, pH 7.5 at 37 °C. After 20 h of incubation, the reactions were stopped by boiling the mixture for 5 min. The resultant suspension was treated with mutanolysin for 20 h at 37 °C. The reactions were stopped by boiling the mixture for 5 min and reduction of the reducing ends by sodium borohydride. After centrifugation (17,000 g for 10 min), the supernatants of the resultant mixtures were subjected to LC/MS analysis, whose conditions were reported previously (Lee et al., 2017).

2.2.3 Protein expression and purification

The gene encoding residues Glu21–Gln273 Pgp3 from *C. jejuni* (strain ATCC

[2] Seven synthetic substrates were synthesized by Mijoon Lee and Dusan Hesek.

[3] LC/MS was done by Mijoon Lee.

33560) was cloned into the expression vector pET-21a (+) (Novagen) using NdeI and XhoI restriction enzymes, producing the recombinant Pgp3 protein with the C-terminal His_6-tag (LEHHHHHH). The Pgp3 protein was expressed in *E. coli* Rosetta2 (DE3) cells (Novagen) induced with 0.5 mM IPTG at 30 °C for 16 h following growth to mid-log phase at 37 °C. The harvested cell pellet was resuspended in a lysis buffer containing 20 mM Tris-HCl pH 7.9, 500 mM sodium chloride, 5 mM imidazole, 10% glycerol, and 1 mM PMSF and was lysed by sonication. After centrifugation at 36,000 g for 1 h at 4 °C, the cell debris was discarded, and the supernatant was applied to an affinity chromatography column of HiTrap chelating HP (GE Healthcare), which was previously equilibrated with the lysis buffer. The column was washed with lysis buffer containing 25 mM imidazole and was eluted with a linear gradient from 25 to 500 mM imidazole. The protein was eluted at 150−200 mM and further purified by gel filtration on a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare), which was previously equilibrated with 20 mM HEPES pH 7.5, 200 mM NaCl, and 1 mM DTT (Figure 2.2). Peak fractions containing the Pgp3 protein were pooled and concentrated to 10 mg ml^{-1} for crystallization. The mutants of Pgp3 (H216A and H247A) were generated by site-directed mutagenesis using primer sets, Pgp3-H216A-F, Pgp3-H216A-R, Pgp3-H247A-F, and Pgp3-H247A-R (Table 2.1) and were purified as described for wild-type Pgp3 (Figure 2.3 and 2.4).

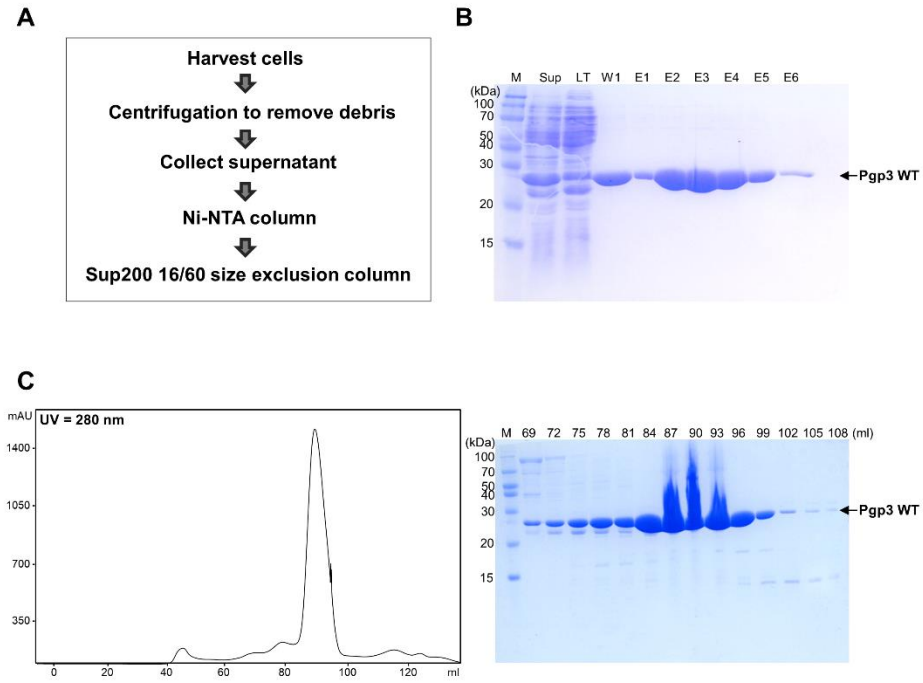


Figure 2.2 Purification of recombinant *C. jejuni* wild-type of Pgp3$_{21\text{-}273}$.

(A) A schematic flow chart displays all purification steps of recombinant *C. jejuni* wild-type of Pgp3$_{21\text{-}273}$. (B) The SDS-PAGE result of Pgp3$_{21\text{-}273}$ after purifying by Ni-NTA affinity chromatography. Purified Pgp3$_{21\text{-}}$ was separated by 12% SDS-PAGE and stained using Coomassie brilliant blue. The protein marker is shown in lane M. The arrow indicates wild-type Pgp3$_{21\text{-}273}$. (C) A FPLC elution profile and SDS-PAGE result of Pgp3$_{21\text{-}273}$ purified by Superdex 200 16/60 size exclusion chromatography.

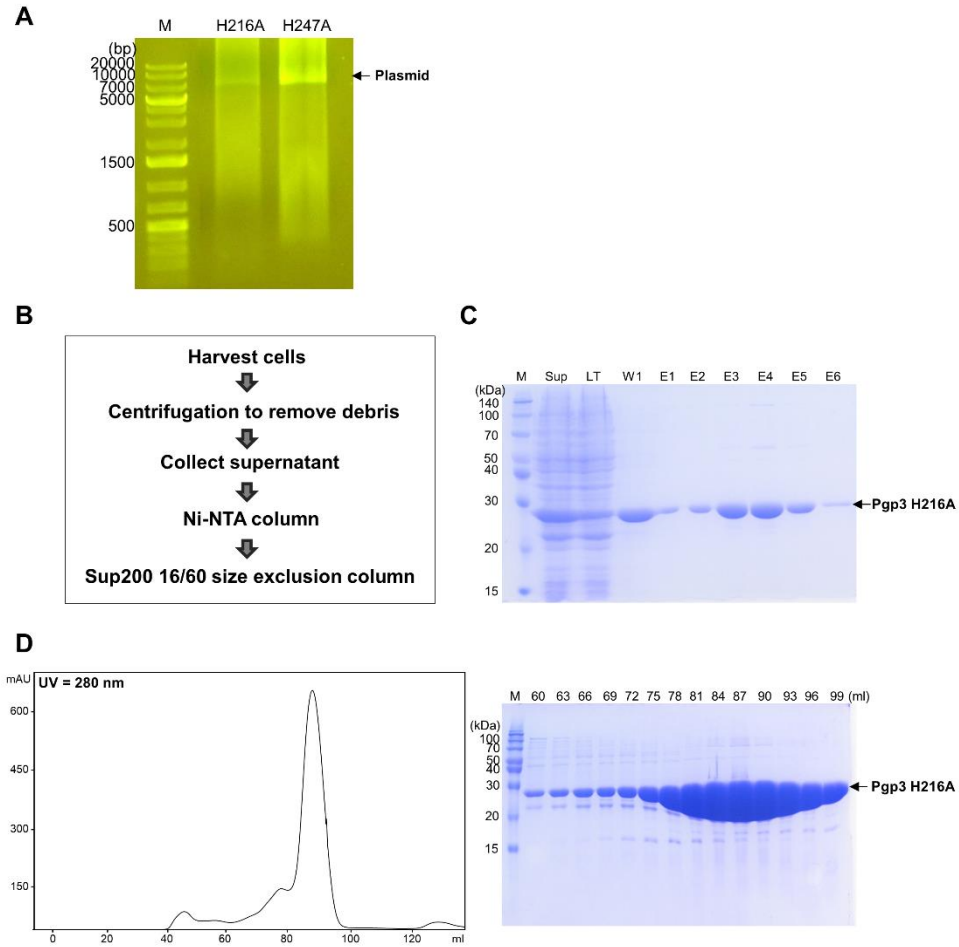


Figure 2.3 Purification of Pgp3$_{21\text{-}273}$ H216A protein from *C. jejuni*.

(A) A site-directed mutagenesis result to generate H216A and H247A variants of Pgp3 is shown. The arrow indicates a plasmid for each variant. (B) A schematic flow chart shows all purification steps of recombinant *C. jejuni* H216A Pgp3$_{21\text{-}273}$ protein. (C) The SDS-PAGE result of H216A Pgp3$_{21\text{-}273}$ after purifying by Ni-NTA affinity chromatography. Purified H216A Pgp3$_{21\text{-}273}$ was separated by 12% SDS-PAGE and stained using Coomassie brilliant blue. The protein marker is shown in lane M. The arrow indicates H216A Pgp3$_{21\text{-}273}$. (D) A FPLC elution profile and SDS-PAGE result of H216A Pgp3$_{21\text{-}273}$ purified by Superdex 200 16/60 size exclusion chromatography.

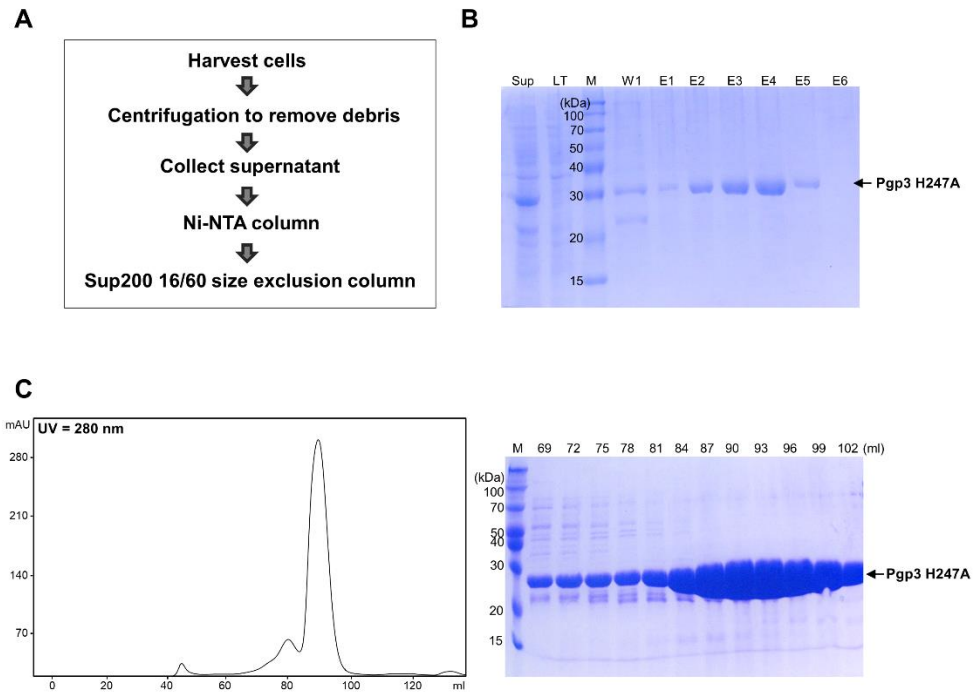


Figure 2.4 Purification of Pgp3$_{21\text{-}273}$ H247A protein from *C. jejuni*.

(A) A schematic flow chart exhibits all purification steps of recombinant *C. jejuni* H247A Pgp3$_{21\text{-}273}$ (B) The SDS-PAGE result of H247A Pgp3$_{21\text{-}273}$ after purifying by Ni-NTA affinity chromatography. Purified H247A Pgp3$_{21\text{-}273}$ was separated by 12% SDS-PAGE and stained using Coomassie brilliant blue. The protein marker is shown in lane M. The arrow indicates H247A Pgp3$_{21\text{-}273}$. (C) A FPLC elution profile and SDS-PAGE result of H247A Pgp3$_{21\text{-}273}$ purified via Superdex 200 16/60 size exclusion chromatography.

2.2.4 Crystallization and data collection[4]

The two structures of wild-type Pgp3 were determined, i.e., tartrate- and citrate-bound forms. Crystals were grown at 296 K by the sitting-drop vapor-diffusion method using the Mosquito robotic system (TTP Labtech). For the tartrate-bound form, the sitting drop was prepared by mixing equal volumes of the protein solution (10 mg ml^{-1}, 0.2 μL) and reservoir solution (0.2 μL) containing 200 mM potassium sodium tartrate, 100 mM tri-sodium citrate at pH 5.6, and 2 M ammonium sulfate. Crystals of tartrate-bound Pgp3 were cryoprotected in the reservoir solution supplemented with 20% (v/v) glycerol and were flash-frozen in a nitrogen gas stream at 100 K. For the citrate-bound form, the sitting drop was prepared by mixing equal volumes of the protein solution (10 mg ml^{-1}, 0.2 μL) and reservoir solution (0.2 μL) containing 10% 2-propanol, 100 mM citrate at pH 5.5, and 20% PEG4000. Crystals of citrate-bound Pgp3 were cryoprotected in the reservoir solution supplemented with 10% (v/v) glycerol and were flash-frozen in a nitrogen gas stream at 100 K. Native data for the tartrate-bound and citrate-bound Pgp3 were collected at 1.86 Å and 1.66 Å resolutions, respectively, using the ADSC Q315r CCD detector at the beamline BL-5C of Pohang Light Source (PLS), Pohang, Korea (Table 2.2).

Crystals of H216A variant in apo-form and tartrate-bound form were grown at 296 K by sitting-drop vapor diffusion method using the Mosquito robotic system (TTP Labtech). For the crystal of the H216A apo-form, sitting drop was prepared by mixing equal volumes of the protein solution (10 mg ml^{-1}, 0.2 μL) and reservoir solution (0.2 μL) containing 200 mM potassium sodium tartrate tetrahydrate, 100 mM sodium citrate tribasic dehydrate at pH 5.6, and 2.0 M ammonium sulfate. The crystal of the H216A tartrate-bound form was

[4] Pentapeptide was synthesized by Jisu Park.

prepared by mixing equal volumes of the protein solution (10 mg ml^{-1}, 0.2 µL) and reservoir solution (0.2 µL) containing 1.1 M malonic acid, 150 mM ammonium citrate tribasic, 72 mM succinic acid, 180 mM dL-malic acid, 240 mM sodium acetate, 300 mM sodium formate, and 96 mM ammonium tartrate dibasic at pH 7.0. Both crystals were soaked in Paratone-N (Hampton Research, Aliso Viejo, CA, USA) before being flash-frozen in a nitrogen stream at 100 K. Native data were collected to 2.10 Å and 2.04 Å resolutions for the apo-form and tartrate-bound form, respectively, using the beamline BL-5C of PLS (Table 2.2).

For the Pgp3 H247A variant crystal, sitting drop was prepared by mixing equal volumes of the protein solution (10 mg ml^{-1}, 0.2 µL) and reservoir solution (0.2 µL) containing 100 mM sodium citrate tribasic dihydrate at pH 4.5, 200 mM potassium sodium tartrate tetrahydrate, and 2.0 M ammonium sulfate. For the penta or tetra-tri peptide complex, H247A protein (10 mg ml^{-1}) in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, and 70 mM peptides in 150 mM Tris-HCl pH 8.0 were mixed in a 1:10 molar ratio and incubated at 277 K for 1 h. Crystals were grown at 296 K by the sitting-drop vapor diffusion method by mixing equal volumes of the protein-peptide mixture (0.2 µL) and reservoir solution (0.2 µL) containing 1.8 M ammonium citrate tribasic at pH 7.0. Crystals reached maximum size within 2-3 days. They were cryoprotected in the reservoir solution supplemented with 30% glycerol and were flash-frozen in a nitrogen gas stream at 100 K. X-ray diffraction data were collected in 1° oscillations at beamline BL-7A of PLS (Table 2.2). Raw X-ray diffraction data for all crystals mentioned above were processed and scaled using the program suit HKL2000 (Otwinowski and Minor, 1997). Data collection statistics are summarized in Table 2.2.

2.2.5 Structure determination and refinement

Two wild-type structures of tartrate- and citrate-bound Pgp3 were determined by molecular replacement utilizing the program MOLREP, with putative peptidase M23 from *Pseudomonas aeruginosa* (PDB code 2HSI) as a search model (unpublished data) (Vagin and Teplyakov, 2010). The peptidase M23 (residues Arg68–Gln281) shows 34% sequence identity with Pgp3 (residues Glu57–Asn269). Manual model building was performed using the program COOT and models were refined with the program REFMAC5, including the bulk solvent correction (Emsley et al., 2010; Murshudov et al., 1997). In total, 5% of the data was randomly set aside as test data for calculating R_{free} (Brunger, 1992). The stereochemistry of the refined models was assessed using MolProbity (Chen et al., 2010). Crystallographic and refinement statistics are summarized in Table 2.2.

The structures of the H216A, H247A, and H247A variants in complex with two peptides were also solved by the molecular replacement method using the monomer model (chain A) of wild-type Pgp3. Subsequent manual model building was carried out using the COOT program and restrained refinement was performed using the REFMAC5 program (Murshudov et al., 1997). Several rounds of model building, simulated annealing, positional refinement, and individual B-factor refinement were performed. Table 2.2 lists the refinement statistics. Atomic coordinates and structure factors of eight crystal structures of Pgp3: (i) tartrate-bound form, (ii) citrate-bound form (two structures), (iii) H216A variant, (iv) H216A variant with tartrate, (v) H247A variant, (vi) H247A variant with pentapeptide (L-Ala^1-γ-D-Glu^2-mDAP^3-D-Ala^4-D-Ala^5), (vii) H247A variant with cross-linked muramyl tetra-tri peptide (NAM-L-Ala^1-γ-D-Glu^2-mDAP^3-D-Ala^4-mDAP^3-γ-D-Glu^2-Ala^1) have been deposited in PDB (PDB ID codes 6JMX, 6JMY, 6KV1, 6JN8, 6JN7, 6JMZ, 6JN1, and 6JN0).

Table 2.1. Primers for Pgp3 structure study.

Primers	Sequence (5´-3´)
Pgp3-NdeI-F	GCGCTCATATGATGAAAGCCTTATGGCTTTTTTTAAG
Pgp3-XhoI-R	GCGCTCTCGAGTCATTGAAAAATAGCATTAAATTTGGAT
Pgp3-H216A-F	GAATTTATTCACAATATTATGCTCTTTCTAAAATCGATGTTAA
Pgp3-h216A-R	TTAACATCGATTTTAGAAAGAGCATAATATTGTGAATAAATTC
Pgp3-H247A-F	GTGGTAGGGTAAGTGGGCCGGCTTTGCATTTTGGAATTTTAGC
Pgp3-H247A-R	GCTAAAATTCCAAAATGCAAAGCCGGCCCACTTACCCTACCAC

Table 2.2 Data collection and refinement statistics.

Data set	wild-type apo open form	wild-type apo closed form	H216A apo open form	H216A apo closed form	H247A apo open form	H247A tetra-tri peptide bound	H247A penta-peptide bound	wild-type apo closed form (Zn SAD)
A. Data collection								
X-ray source	PLS-5C	PLS-5C	PLS-5C	PLS-5C	PLS-7A	PLS-7A	PLS-7A	PLS-7A
X-ray wavelength (Å)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.2824
Space group	$P6_1$	$P3_221$	$P6_1$	$P2_1$	$P6_1$	$P6_1$	$P6_1$	$P3_221$
Unit cell length (a, b, c, Å)	114.5, 114.5, 57.9	57.9, 57.9, 152.7	113.2, 113.2, 54.5	82.3, 105.6, 86.2	114.3, 114.3, 55.6	114.8, 114.8, 55.6	114.3, 114.3, 55.9	58.1, 58.1, 152.5
Unit cell angle (α, β, γ, °)	90, 90, 120	90, 90, 120	90, 90, 120	90, 107.2, 90	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution range (Å)	50.0–1.86 (1.89–1.86)[a]	50.0–1.66 (1.69–1.66)[a]	50.0–2.10 (2.14–2.10)[a]	50.0–2.04 (2.08–2.04)[a]	50.0–1.92 (1.95–1.92)[a]	50.0–2.20 (2.24–2.20)[a]	50.0–2.40 (2.44–2.40)[a]	50.0–1.72 (1.75– 1.72)[a]
Total/unique reflections	804,553/36,515	481,855/ 35,949	501,309/ 23,195	622,169/ 89,706	619,581/ 31,366	425,878/ 22,451	62,308/ 16,610	1,141,822/ 32,441
Completeness (%)	100.0 (100.0)[a]	100.0 (100.0)[a]	99.8 (100.0)[a]	99.8 (99.1)[a]	98.7 (99.2)[a]	99.9 (100)[a]	99.0 (100)[a]	99.9 (100)[a]
Average I/σ (I)	78.3 (5.6)[a]	51.1 (4.9)[a]	75.2 (8.4)[a]	42.3 (5.2)[a]	73.7 (9.0)[a]	74.6 (14.4)[a]	23.8 (2.1)[a]	102.7 (11.4)[a]
R_{merge}[b] (%)	7.7 (90.0)[a]	6.2 (45.6)[a]	13.4 (84.9)[a]	10.4 (69.0)[a]	10.0 (93.4)[a]	12.3 (96.1)[a]	7.2 (79.4)[a]	12.0 (68.1)[a]
B. Model refinement statistics								
Resolution range (Å)	50.0–1.86	50.0–1.66	50.0–2.10	50.0–2.04	50.0–1.92	50.0–2.20	50.0–2.40	50.0–1.72
R_{work} / R_{free}[c] (%)	17.2 / 19.9	17.3 / 20.6	18.2 / 21.3	16.7 / 19.4	17.1 / 19.7	16.1 / 19.9	17.8 / 22.6	21.7 / 26.0
Monomers per asymmetric unit	1	1	1	3	1	1	1	1

Number of non-hydrogen atoms / average *B*-factor (Å^2)								
Protein	2,028/ 43.8	2,038 / 31.8	2,038 / 51.5	6,249 / 35.1	2,014 / 41.4	2,014 / 46.8	2,006 / 60.5	2,038 / 32.9
Water oxygen	242 / 50.3	210 / 41.8	136 / 52.0	651 / 44.2	135 / 47.3	157 / 54.6	99 / 60.8	218 / 41.8
Zn^{2+}	1 / 33.0	1 / 21.9	1 / 54.0	3 / 26.1	1 / 39.8	1 / 53.7	1 / 83.0	1 / 34.0
Tartrate	10 / 63.3	-	-	30 / 51.2				
Citrate	-	13 / 34.5						13 / 35.2
Glycerol	18 / 58.8	-						
Peptide					-	30 / 77.3	23 / 85.0	
Sulfate			5 / 72.2	-				
R.m.s. deviations from ideal geometry								
Bond lengths (Å)/ bond angles (°)	0.007 / 0.80	0.006 / 0.78	0.008 / 0.89	0.007 / 0.84	0.006 / 0.79	0.007 / 0.91	0.008 / 1.09	0.006 / 0.79
Protein-geometry analysis								
Ramachandran favored (%)	96.5	95.3	92.7	95.5	93.3	94.9	95.6	95.3
Ramachandran allowed (%)	3.5	3.9	5.0	3.9	5.5	4.7	3.2	4.3
Ramachandran outliers (%)	0	0.8	2.3	0.6	1.2	0.4	1.2	0.4

Footnotes for Table 2.2

[a]Values in parentheses refer to the highest resolution shell.

[b]$R_{merge} = \Sigma_{hkl}\Sigma_i \mid I_i(hkl) - \langle I(hkl)\rangle \mid / \Sigma_{hkl}\Sigma_i\, I_i(hkl)_i$, where $I(hkl)$ is the intensity of reflection *hkl*, Σ_{hkl} is the sum over all reflections, and Σ_i is the sum over i measurements of reflection *hkl*.

[c]$R = \Sigma_{hkl} \mid |F_{obs}| - |F_{calc}| \mid / \Sigma_{hkl} |F_{obs}|$, where R_{free} was calculated for a randomly chosen 5% of reflections, which were not used for structure refinement and R_{work} was calculated for the remaining.

2.2.6 SEC-MALS

SEC-MALS experiments for wild-type and mutants (H216A and H247A) of Pgp3 were performed using an FPLC system (GE Healthcare) connected to a Wyatt MiniDAWN TREOS MALS instrument and a Wyatt Optilab rEX differential refractometer. A Superdex 200 10/300 GL (GE Healthcare) gel filtration column pre-equilibrated with buffer containing 20 mM Tris-HCl pH 8.0 and 200 mM NaCl, was normalized using ovalbumin. Proteins were injected at a flow rate of 0.4 ml min^{-1}. Data were analyzed using the Zimm model for static light-scattering data fitting and graphs were constructed using EASI graph with a UV peak in the ASTRA V software (Wyatt).

2.2.7 Computational modeling[5]

We tested the conformational transition between two structures of wild-type Pgp3 upon substrate binding using computational methods. Three-dimensional crystal structures of wild-type and mutant Pgp3 (H216A and H247A) were superimposed in PyMol v1.7 and the coordinates of the co-crystallized ligand in H247A mutant, i.e. tetra-tri peptide ligand, were imported inside wild-type Pgp3. The ligand (mDAP^3-D-Ala^4-mDAP^3) position was used to append the peptide chains (L-Ala^1-γ-D-Glu^2 and γ-D-Glu^2-L-Ala^1) on either ends to complete the model of cross-linked peptide chain (L-Ala^1-γ-D-Glu^2-mDAP^3-D-Ala^4-mDAP^3-γ-D-Glu^2-L-Ala^1). In the lack of concrete evidence about the nature of Pgp3 substrate (murein sacculus or muropeptides) and considering the relevance of AmiA in *C. jejuni*, the glycan residues (NAG-NAM-NAG) were attached only on one end of peptide chain in order to leave one side open to facilitate access of Pgp3. Ultimately, as mentioned a priori, the glycan residues are solvent exposed and impact peptide-binding minimally. Therefore, even our general model can

[5] Computational modeling was done by Neha Rana.

be fitted to all plausible Pgp3 substrates (murein sacculus that are modified by AmiA or NAG-NAM, NAG-anhydroNAM, NAM muropeptides on either or both ends). For modeling wild-type complex, the active-site Zn^{2+} and its tetrahedrally coordinated residues, His168, Asp172, His249 and peptidoglycan substrate were treated with QM/MM. DFT level of theory with B3LYP functional and 6-31G* basis set was used to optimize geometry of residues and perform charge refitting with RESP. A bonded-coordination model between Zn^{2+}, NE atom of His168, OD1 atom of Asp172, ND atom of His249 and polarized carbonyl oxygen of substrate's -D-Ala^4-$mDAP^3$- amide bond was constructed using MCPB module of AmberTools[17]. Additionally, His216 and His247 were restrained along with a water molecule in Wat-1 location, unless otherwise specified. The wild-type complex files were prepared with Amber forcefield ff14SB and waters were treated with either TIP4P or OPC models inside a rectangular box with 10 Å buffer. All MD simulations were carried out using GPU-supported Amber18 MD package. The equilibrium dynamics of wild-type complex was conducted with constant temperature and pressure (NPT) and periodic boundary conditions. Particle Ewald Mesh (PME) was used to calculate long-range electrostatic interactions and the non-bonded cutoff was set to 10.0 Å. SHAKE was applied to constrain bonds involving hydrogen and temperature was maintained at 300 K using Langevin thermostat with a collision frequency of 5.0 ps^{-1}. An average pressure of 1 atm was maintained with isotropic scaling and 2 ps relaxation time. The integration time step of 2 fs was used, unless otherwise specified. The complex was first minimized, and the temperature was raised from 0 K to 300 K in 40000 steps gradually. A brief equilibration of 500 ps was performed before starting production run of 500 ns. Three additional 10 ns apo-wild-type MD simulations, (i) restrained protein, unrestrained water, (ii) temperature Replica-exchange MD (T-REMD), and (iii) targeted MD (TMD) were also performed to study the apo-state Zn^{2+} coordination character and

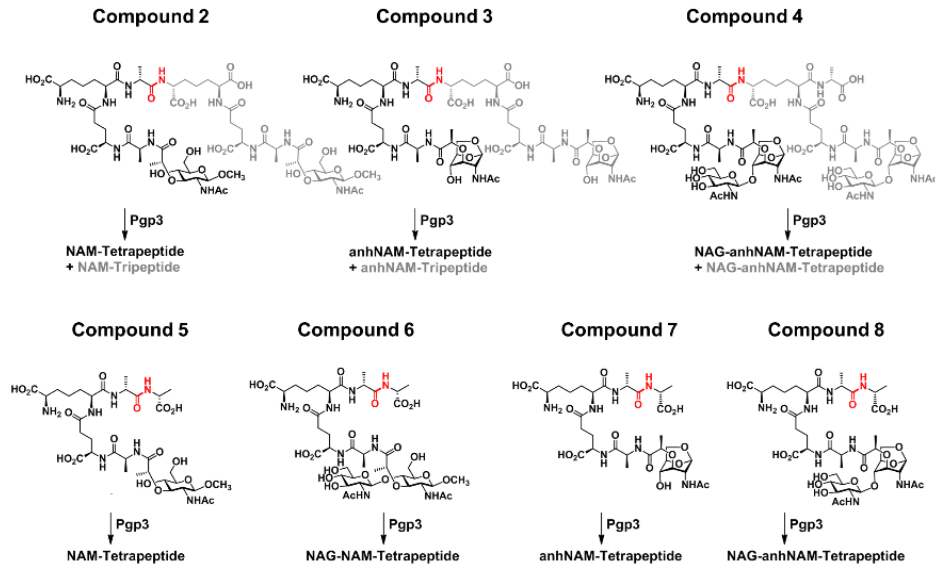
conformational dynamics between open and closed states, respectively.

In REMD, 60 replicas were selected depending on the square root of the number of atoms in the system and the temperature was varied between 270 K and 600 K in geometric progression. The chirality restraints were imposed on the backbone to avoid non-physical rotation of peptide bonds at high temperatures. The REMD was carried out using constant temperature, constant volume (NVT) dynamics with 2 fs time step and exchange between replicas was allowed every 1000 steps. In TMD, an additional energy term was added based on the RMSD between L1 loop residues, in the open and closed apo-states of wild-type Pgp3. The energy term has a harmonic nature; $V = ½ \times k \times (RMSD(t) - RMSD_0(t))^2$, where k has a force constant value of 5 kcal mol^{-1} $Å^2$ (calibrated after several trials), RMSD(t) is the RMSD at simulation time t between current snapshot and reference structure and $RMSD_0(t)$ is the prescribed RMSD at simulation time t between current snapshot and reference structure. The starting RMSD was set to 6.983 Å and then it was decreased gradually to 0 Å in 5 ns and kept at 0 Å for next 5 ns. All the clustering treatments were done by CPPTRAJ in AmberTools[17] using DBSCAN clustering method.

2.3 Results

2.3.1 Pgp3 has D,D-carboxy- and D,D-endopeptidase activities

In order to investigate the role of Pgp3 in peptidoglycan turnover, we performed assays using seven synthetic muramyl peptide substrates (compounds 2-8, Figure 2.5), each prepared in multistep syntheses by known procedures (Lee et al., 2015; Lee et al., 2010; Hasek et al., 2009). Pgp3 exhibited both D,D-carboxypeptidase and D,D-endopeptidase activities based on LC/MS analysis (Figures 2.5 and 2.6). We were not able to isolate the sacculus of *C. jejuni* in quantity. However, the peptidase activity was also documented with the *E. coli* sacculus as its macromolecular substrate (Figure 2.7). The mature *E. coli* sacculus is devoid of the full-length pentapeptide, so the D,D-carboxypeptide activity could not be monitored, but we documented the D,D-endopeptidase activity (Figure 2.7). Increasing the concentration of Pgp3 diminishes the D,D-crosslinked peptide (colored in blue, green and red in Figure 2.7).



Peptidase activities of Pgp3		
D,D-Endopeptidase activity	2	70%
	3	95%
	4	100%
D,D-Carboxypeptidase activity	5	100%
	6	87%
	7	100%
	8	100%

Figure 2.5 Synthetic peptidoglycan substrates and reaction products.

Chemical structures of the synthetic peptidoglycan substrates used in this study and their corresponding Pgp3 reaction product. Percentage of product formation for a given substrate.

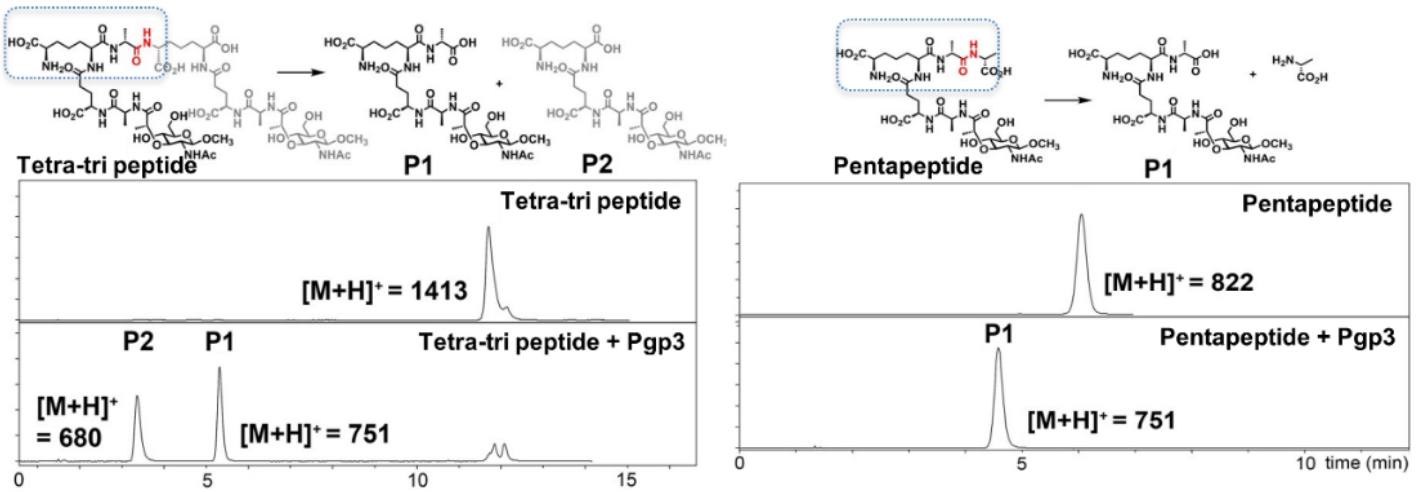
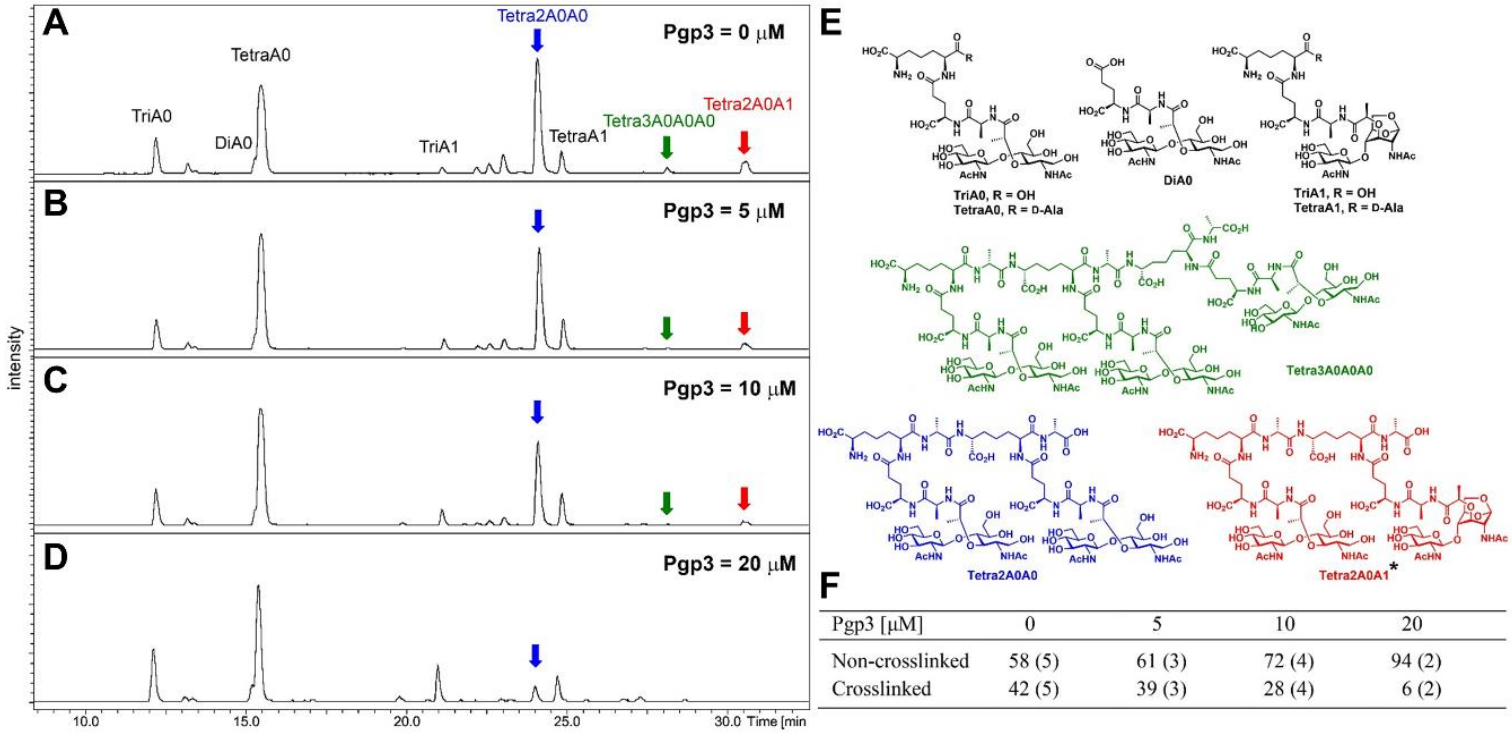


Figure 2.6 LC-MS traces of D,D-endopeptidase and D,D-carboxypeptidase activities of Pgp3.

LC-MS traces of D,D-endopeptidase and D,D-carboxypeptidase activities of Pgp3 with tetra-tri peptide and pentapeptide, respectively. Acceptor strands of crosslinked peptides are shown in gray. The shared structural component between the two substrates is in blue dotted box.



Pgp3 [μM]	0	5	10	20
Non-crosslinked	58 (5)	61 (3)	72 (4)	94 (2)
Crosslinked	42 (5)	39 (3)	28 (4)	6 (2)

Figure 2.7 The reaction of sacculus with Pgp3, followed by mutanolysin.

(A-D) LC-MS traces of the reactions with varying concentration of Pgp3. (E) Chemical structures of major products from the reactions. (F) Summary of product analysis of the reactions (amounts are expressed as a percentage of the total extracted-ion chromatogram peak area;

average values of three runs with error in parenthesis). *The position of NAG-anhNAM was not determined (panel E). The given structure is one of two possible isomers. Colored structures (in blue, green and red) in panel e are muropeptides containing D,D-crosslinks, substrates for Pgp3. Each is shown by correspondingly colored arrows. The fate of those substrates is shown in panels A-D. As the concentration of the enzyme is increased, less substrate survives.

2.3.2 Crystal structure of Pgp3 with open and closed conformation

In order to gain mechanistic insight into how Pgp3 works, we have determined the X-ray structures of the wild-type Pgp3 in two crystal forms (bound separately to tartrate and to citrate) (Figures 2.8 and 2.9). In both structures, the crystallographic asymmetric unit contained two Pgp3 monomers (chain A and chain B) with space groups of $P6_1$ and $P3_221$, respectively (Table 2.2). For the tartrate-bound wild-type, a model accounting for all the amino acids of the recombinant Pgp3 (Glu21-Asn273) was refined at a resolution of 1.86 Å to R_{work} and R_{free} values of 17.2% and 19.9%, respectively (Table 2.2). The citrate-bound wild-type was refined at a resolution of 1.66 Å to R_{work} and R_{free} values of 17.3% and 20.6%, respectively (Table 2.2). When the structure of Pgp3 in complex with tartrate was overlaid with that of the citrate complex, their overall structures were similar with RMSD of 2.95 Å over 256 equivalent Cα positions (Figure 2.9). However, both structures showed a large deviation (RMSD 6.98 Å) in two regions (residues Glu99-Phe109 and Thr159-Val164) near the active-site (Figure 2.9B). This conformational flexibility is discussed later and has mechanistic implications. We designated the tartrate-bound complex as the open form and the citrate-bound Pgp3 as the closed form, based on the solvent exposure of the active site.

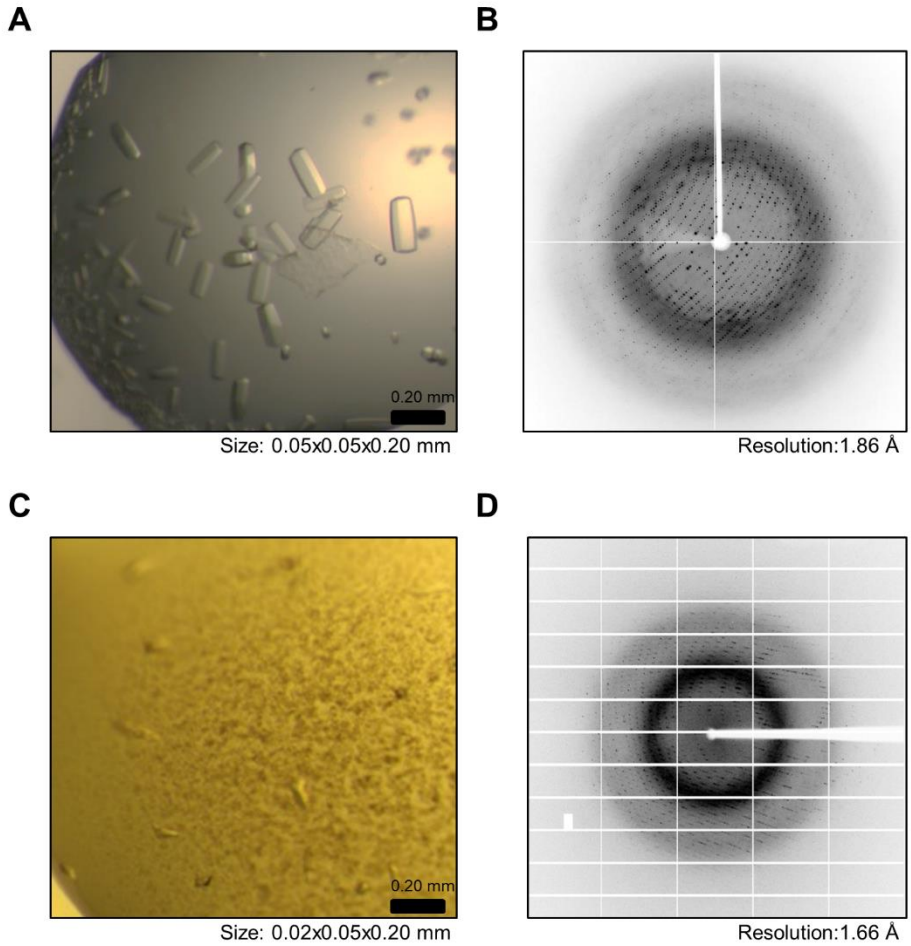


Figure 2.8 Crystal pictures and diffraction patterns of wild-type Pgp3.

(A) A crystal picture of wild-type Pgp3 (opened form) is shown. The approximate dimensions of the crystal are 0.05 mm × 0.05 mm × 0.20 mm (Width × Depth × Height). (B) The X-ray diffraction pattern of wild-type Pgp3 is represented. (C) A crystal picture of wild-type Pgp3 (closed form) is shown. The approximate dimensions of the crystal are 0.02 mm × 0.05 mm × 0.20 mm (Width × Depth × Height). (D) The X-ray diffraction pattern of wild-type Pgp3 is represented.

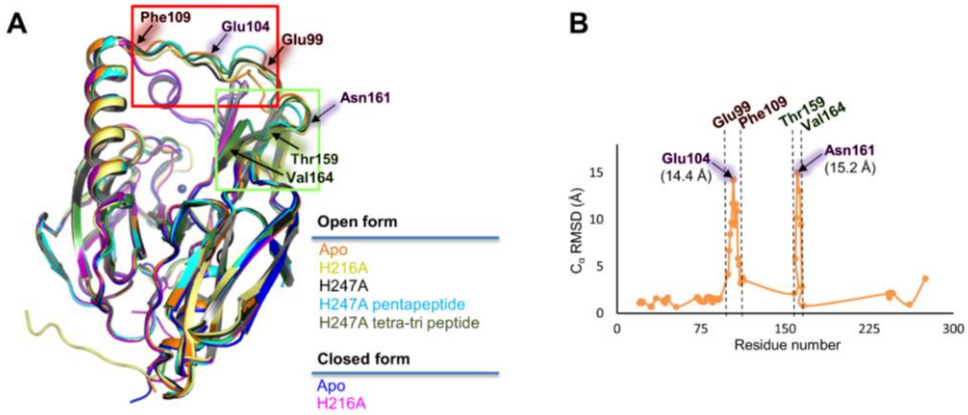


Figure 2.9 Overall structures of Pgp3 protein.

(A) Structural comparisons of seven crystal structures of Pgp3 with open and closed conformations. The superposition of seven crystal structures of Pgp3 (wild-type open form in orange, H216A mutant open form in light yellow, H247A mutant in black, H247A mutant with pentapeptide in light blue, H247A mutant with tetra-tri peptide in dark green, wild-type closed form in blue, and H216A mutant closed form in pink). The red and green boxes show the most deviation regions in the distance plots. The red box includes those from the Glu99 to Phe109 residues. The green box covers those from the Thr159 to Val164 residues. The most deviated residues between open and closed structures are Glu104 and Asn161 which are indicated as purple arrows in the figure. (B) Distance plots between wild-type Pgp3 in the open form (PDB ID: 6JMX) and corresponding Cα positions of wild-type Pgp3 in the closed form (PDB ID: 6JMY).

2.3.3 Domain architecture and structural comparisons

A search for overall structural similarities with the full-length Pgp3 using the program DALI failed to reveal any significant matches (Holm et al., 1993). For more detailed analysis, we examined each domain individually. The Pgp3 monomer is structurally divided into four regions: the N-terminal domain (NTD, residues 21-95), a flexible linker (residues 96-147), the LytM domain (residues 148-259), and the C-terminal helix (residues 260-273) (Figure 2.10A). The strong electron density observed at the metal-binding site of LytM domain was confirmed to be the Zn^{2+} ion by calculating anomalous difference maps using data collected at the zinc absorption edge (1.2824 Å) (Figure 2.10B). The Pgp3 LytM domain contains a highly conserved tetrahedral Zn^{2+}-ion-binding motif (H^{168}xxxD^{172} and HxH^{249}). The NTD comprises six β-strands arranged in two β-sheets (β2↓-β4↑ and β1↑-β6↑-β5↓-β3↑) (Figure 2.10A). According to the DALI search using NTD, the highest similarity was detected in the immunoglobulin domain of mouse leukemia inhibitory factor receptor (mLIFR) (PDB code 2Q7N; chain C; residues 211-286; Z-score 7.2), indicating that the NTD adopts an immunoglobulin-like fold (Holm, et al., 2010; Huyton, et al., 2007). However, the key residues (Lys153, Phe156, and Lys159) for ligand binding in mLIFR were absent in the NTD of Pgp3, suggesting that they are not functionally related (Hudson, et al., 1996). The flexible linker connecting the NTD and LytM domain contains one short β-strand (β7), one α-helix (α1), and one loop (L1). Following the flexible linker, Pgp3 contains the Zn^{2+}-coordinated LytM domain with the canonical fold observed in the M23 peptidase family. The LytM domain of Pgp3 comprises a larger, eight-stranded anti-parallel β-sheet (β11↑-β12↓-β13↑-β14↓-β15↑-β16↓-β17↑-β18↓) and a smaller, two-stranded anti-parallel β-sheet (β8↓-β9↑). An anti-parallel arrangement of two β-strands (β8↓-β9↑) is present along with the β7 strand of the flexible linker (Figure 2.10A). As expected, the DALI search with the LytM domain of Pgp3 reveals high structural

similarity with other LytM domains of peptidoglycan carboxypeptidases (Csd3 from *H. pylori*, PDB code 4RNZ) and the M23 peptidase family (PDB codes 5J1K, 5J1L, and 4ZYB) with high Z-scores (17.8, 16.3, 16.2, and 15.9, respectively) (Holm, et al., 2010). The length of the C-terminal helix (residues 260-273) of Pgp3 is shorter than that of two other orthologues in *H. pylori* Csd1 and Csd3. The overall structure of Pgp3 differs remarkably from other structurally characterized Csd proteins (Figures 2.11 and 2.12).

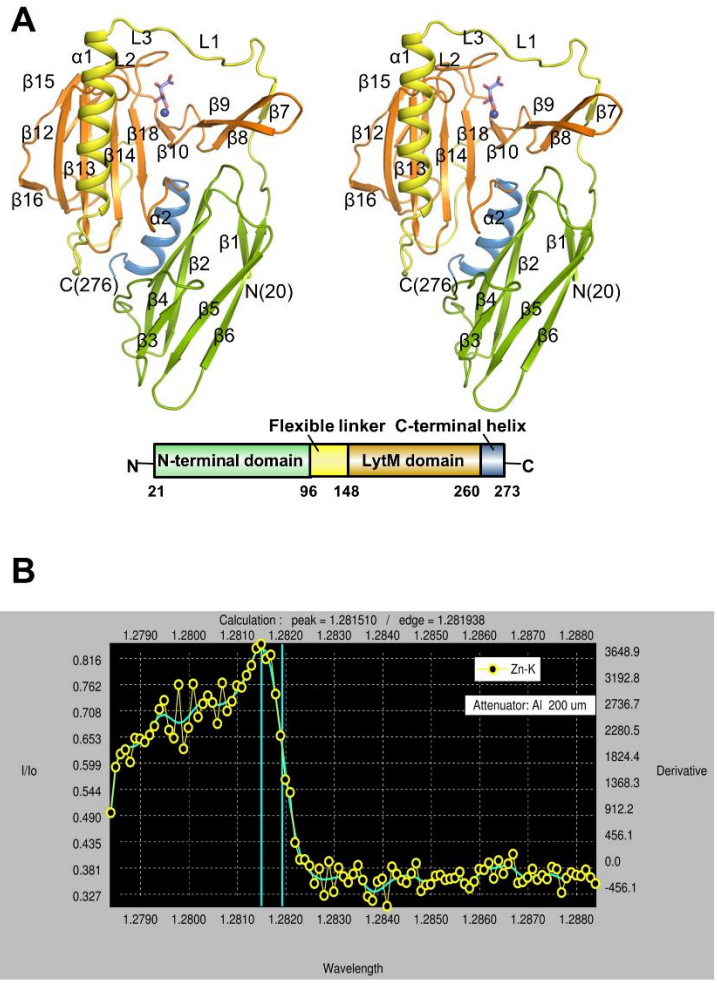


Figure 2.10 Overall structures and structural comparisons of Pgp3

(A) Domain architecture of Pgp3 and its stereo ribbon diagram. The N-terminal domain (NTD) and C-terminal domain (CTD) are shown in light green and blue, respectively, whereas the flexible linker and LytM domain are drawn in yellow and orange, respectively. Tartrate molecule is shown as sticks. (B) X-ray absorption data for Zn^{2+}.

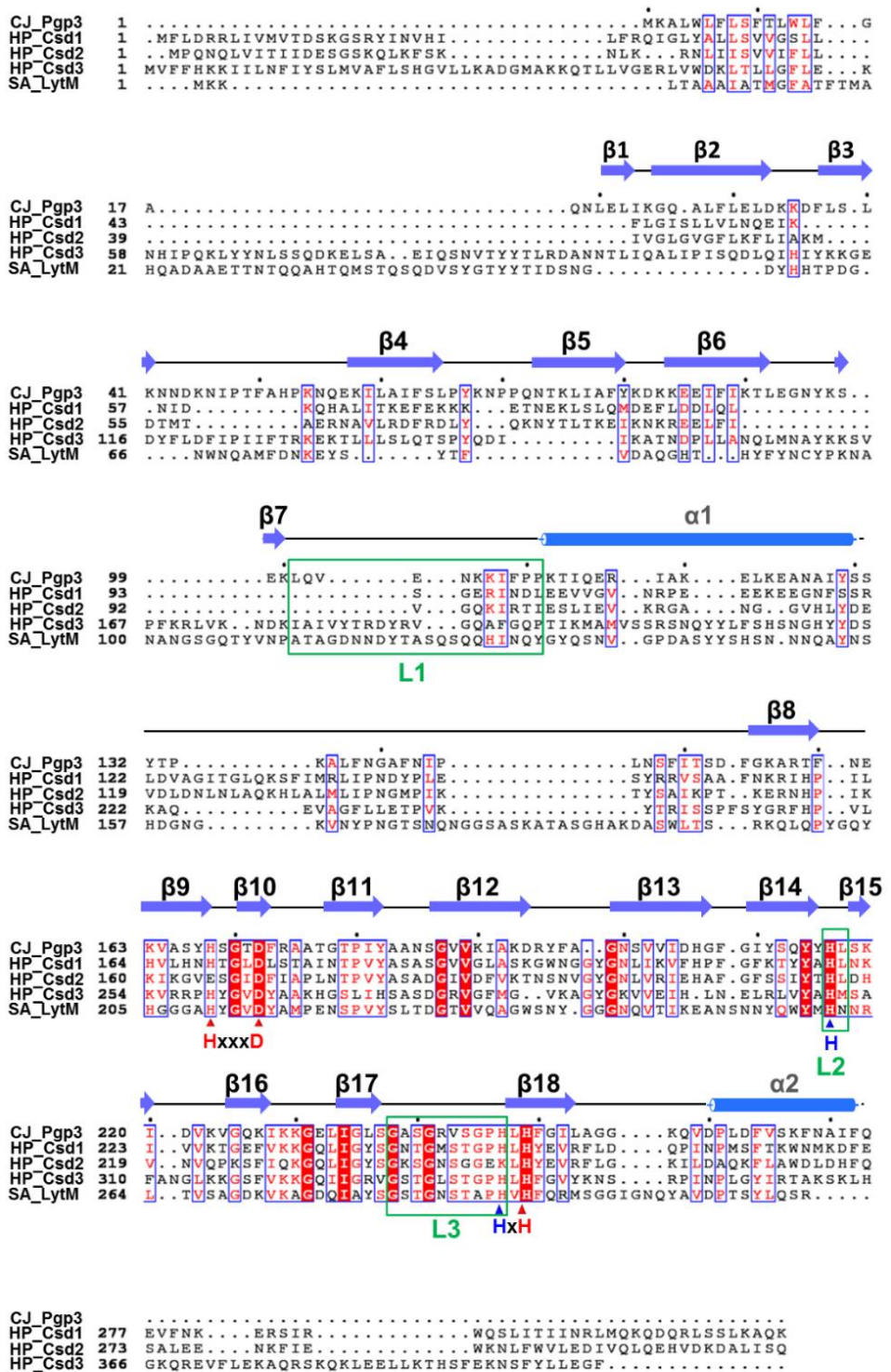


Figure 2.11 Sequence alignment of *C. jejuni* Pgp3 with its homologs.

Multi-alignment of Pgp3 (UniProtKB/Swiss-Prot accession number A8118_01115) against Csd1 from *H. pylori* (UniProtKB/Swiss-Prot accession number HP1543), Csd2 from *H. pylori* (UniProtKB/Swiss-Prot accession number HP1544), Csd3 from *H. pylori* (UniProtKB/Swiss-Prot accession number HP0506), and LytM from *S. aureus* (UniProtKB/Swiss-Prot accession number O33599). Secondary structural elements were assigned using PyMOL and every tenth residue is marked with a black dot. Red triangles indicate the conserved residues in HxxxD and HxH motifs that are important for the metallopeptidase activity. Three loops are indicated by green boxes. Blue triangles indicate the conserved residues for water coordination in substrate hydrolysis activity.

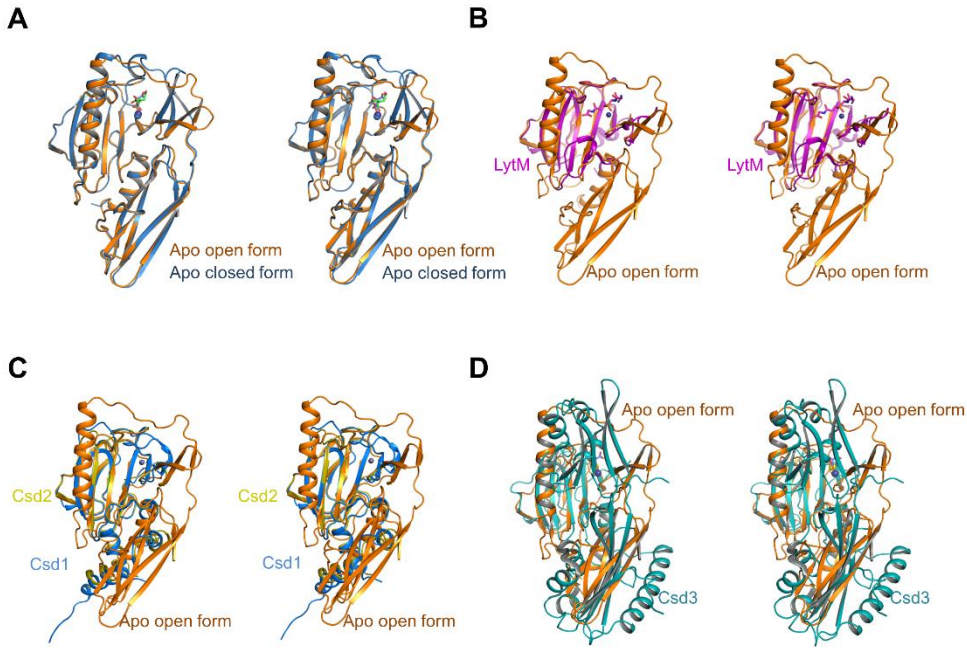


Figure 2.12 Structural comparisons between Pgp3 and M23 metalloprotease.

(A) Structural comparison between open (orange, PDB ID: 6JMX) and closed forms (blue, PDB ID: 6JMX) of Pgp3. (B) Stereo view of the superposition between the LytM domain from *S. aureus* (PDB ID: 4ZYB, magenta) and wild-type Pgp3 (open form, orange, PDB ID: 6JMX). (C) Stereo view of the superposition between Csd1/Csd2 from *H. pylori* (PDB ID: 5J1L and 5J1K), and wild-type Pgp3 (open form, PDB ID: 6JMX), shown in blue, yellow, and orange, respectively. (D) Stereo view of the superposition between Csd3 from *H. pylori* (PDB ID: 4RNZ, dark cyan) and wild-type Pgp3 (open form, orange, PDB ID: 6JMX).

2.3.4 Crystal structure of H247A and H216A variants Pgp3

Crystals of Pgp3 could only be produced in the presence of either tartrate or citrate molecules. Both molecules bound to the active site of Pgp3 via several hydrogen bonds and salt-bridge interactions between the carboxyl groups of tartrate or citrate and the Zn^{2+} ion (Figures 2.13A and 2.13B). We decided to mutate residues interacting with tartrate or citrate in order to abrogate binding of the molecules to the active site (H216A and H247A) (Figures 2.13C and 2.13D). By comparing wild-type Pgp3 with other M23 family proteins (PDB codes 3IT5 and 3IT7), we noticed that the positions of two conserved water molecules (hereafter referred to as Wat-1 and Wat-2) were substituted by tartrate or citrate oxygen atoms (Spencer, et al., 2010). Specifically, in the tartrate-bound complex, Wat-1 and Wat-2 were displaced by tartrate O3 and O41 atoms and in the citrate-bound wild-type Pgp3, they were displaced by citrate O5 and O6 atoms, respectively. Furthermore, in the tartrate-bound complex, the side chain of His216 (NE2 atom) formed a hydrogen bond with the O3 of the tartrate molecule (3.4 Å), and the side chain of His247 (NE2 atom) formed another hydrogen bond with tartrate O3 atom (2.7 Å). In the citrate-bound complex, the side chain of His247 (NE2 atom) formed one hydrogen bond with the citrate O5 atom (2.7 Å), and another hydrogen bond with citrate O7 atom (3.2 Å). Interestingly, Arg158 showed large conformational change, pulling center of mass of β7↑-β8↓-β9↑ sheet by 49° (and 6 Å) in the direction of the active site (compared to the positions in the tartrate-bound complex), and engaged in salt-bridge interactions with the citrate carboxyl group.

The amino acids His216 and His247 are strictly conserved for Wat-1 and Wat-2 coordination in the M23 metallopeptidase family, and for the core of the catalytic machinery. We decided to mutate His216 and His247 individually to alanine in order to abrogate the binding of tartrate or citrate at the active site and impair catalysis too. Indeed, when we solved the crystal structure of the

H247A variant, it showed the absence of citrate or tartrate electron densities within the active site and the protein only assumed the open conformation, which was fortuitous for efforts toward solving the structure of the enzyme complexes with substrate analogs (Figures 2.13C). In the case of the H216A variant, tartrate was sometimes observed in the closed form crystal structure (Figures 2.13D); hence only the H247A variant was pursued further. The crystallographic data parameters are presented in Table 2.2.

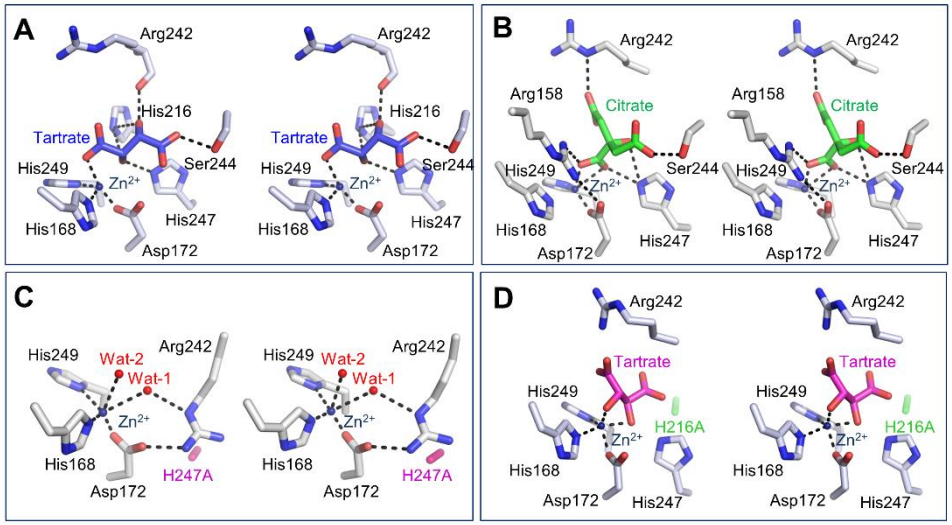


Figure 2.13 Stereo diagrams of Pgp3 active sites.

(A-D) Stereo diagrams of wild-type Pgp3 active sites around (A) tartrate (light blue) and (B) citrate (green) molecules. Tetrahedral coordination of Zn^{2+} and hydrogen bonds is represented by black dotted lines, whereas red spheres represent water molecules. (C) Stereo diagrams of H247A mutant (open form). The position of H247A mutation is shown in pink. (D) Stereo diagrams of H216A mutant in complex with tartrate (magenta). The position of H216A mutation is shown in green.

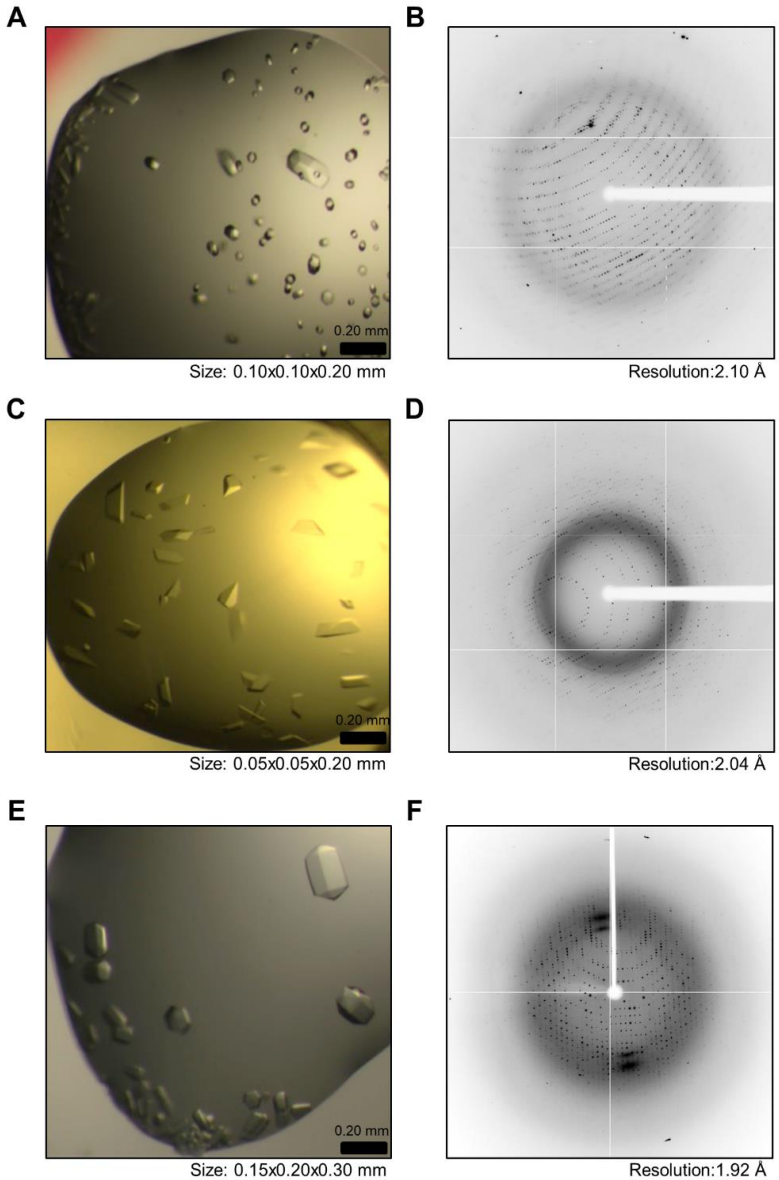


Figure 2.14 Crystal pictures and diffraction patterns of apo-H216A and apo-H247A variants Pgp3.

(A) A crystal picture of H216A Pgp3 (opened form) is shown. The approximate dimensions of the crystal are 0.10 mm × 0.10 mm × 0.20 mm (Width × Depth × Height). (B) The X-ray diffraction pattern of H216A Pgp3 (opened form) is represented. (C) A crystal picture of H216A Pgp3 (closed form) is shown. The approximate dimensions of the crystal are 0.05 mm × 0.05 mm × 0.20 mm

(Width × Depth × Height). (D) The X-ray diffraction pattern of H216A Pgp3 (closed form) is represented. (E) A crystal picture of H247A Pgp3 (opened form) is shown. The approximate dimensions of the crystal are 0.15 mm × 0.20 mm × 0.30 mm (Width × Depth × Height). (F) The X-ray diffraction pattern of H247A Pgp3 (opened form) is represented.

2.3.5 Crystal structure of Pgp3 in complex with its substrates

By using the H247A mutant variant, we successfully solved two structures of Pgp3 in complex with a pentapeptide (1, Ac-L-Ala^1-γ-D-Glu^2-mDAP^3-D-Ala^4–D-Ala^5) and with a cross-linked muramyl tetra-tri peptide (2, NAM-L-Ala^1-γ-D-Glu^2-mDAP^3-D-Ala^4–mDAP^3-γ-D-Glu^2-L-Ala^1-NAM) (Figure 2.15). In the presence of the pentapeptide, excellent electron-density map was obtained for the mDAP^3-D-Ala^4-D-Ala^5 segment (Figure 2.16). In the presence of the crosslinked heptapeptide (tetra-tri peptide), excellent electron density was seen for mDAP^3-D-Ala^4–mDAP^3. The importance of these two complexes is that the scissile bond in the peptide substrate was seen bound within the active site. The segments of the synthetic substrates that were not seen in the structures argue for mobility at those sites.

Both substrates bound the deep cleft of LytM domain with a 1:1 stoichiometry (Figure 2.16). The tetra-tri peptide fragment (mDAP^3-D-Ala^4-mDAP^3) mainly interacted with Pgp3 via its two terminal mDAP moieties. Tyr129, Arg195, and Tyr215 of Pgp3 contributed to the binding by forming several hydrogen-bonding networks. The hydroxyls of Tyr129 and Tyr215 were bound to O7 and O8 atoms of mDAP^3 (from the donor strand) via hydrogen bonds (3.0 Å and 2.5 Å, respectively). The NH1 atom of Arg195 formed a salt bridge with mDAP^3 (from the donor strand). The O5 and N1 atoms of the other mDAP^3 moiety (from the acceptor strand) formed hydrogen bonds with the backbone nitrogen atoms of Val243 and Ser244. These interactions are discussed later in the context of computational simulations with the peptide substrate based on the X-ray structure. When the electrostatic potentials are drawn at the surface of Pgp3, positive electrostatic potentials that are made by the flexible linker (α1 and L1), LytM domain (β8 and β9), and NTD (β4) were widely distributed along the substrate-binding pocket (Figure 2.16). Positive charges at the substrate-binding site are complemented by the negatively

charged peptidoglycan.

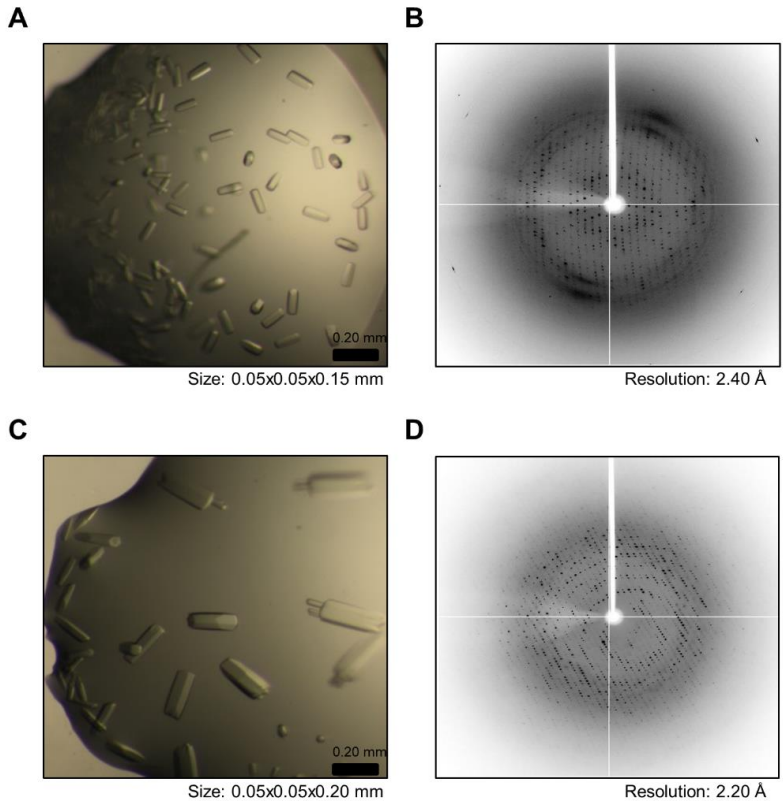


Figure 2.15 Crystal pictures and diffraction patterns of H247A Pgp3 in complex with a pentapeptide and a tetra-tri peptide.

(A) A crystal picture of H247A Pgp3 in complex with a pentapeptide (opened form) is shown. The approximate dimensions of the crystal are 0.05 mm × 0.05 mm × 0.15 mm (Width × Depth × Height). (B) The X-ray diffraction pattern of H247A Pgp3 in complex with a pentapeptide is represented. (C) A crystal picture of H247A Pgp3 in complex with a tetra-tri peptide (opened form) is shown. The approximate dimensions of the crystal are 0.05 mm × 0.05 mm × 0.20 mm (Width × Depth × Height). (D) The X-ray diffraction pattern of H247A Pgp3 in complex with a tetra-tri peptide is represented.

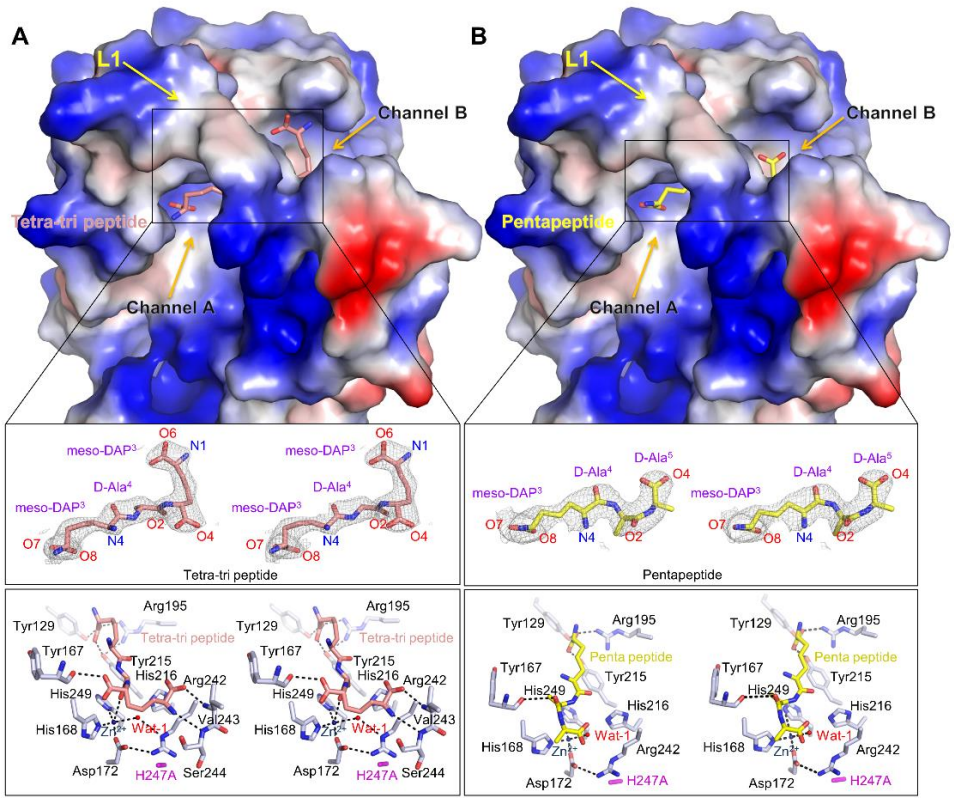


Figure 2.16 Structures of Pgp3 H247A mutant in complex with tetra-tri and pentapeptide.

(A) Surface representation of the electrostatic potential of Pgp3 in complex with tetra-tri peptide. The bound *m*DAP^3-D-Ala^4-*m*DAP^3 moiety is shown in pink sticks for carbons and expanded with the stereo $2F_O$-F_C electron density map (1.0 σ). Tetrahedral coordination of Zn^{2+} and hydrogen bonds are represented by black dotted lines, whereas red spheres represent water molecules. (B) Surface representation of the electrostatic potential of Pgp3 in complex with pentapeptide. The bound *m*DAP^3-D-Ala^4- D-Ala^5 moiety is shown in yellow sticks for carbons and expanded with the stereo $2F_O$-F_C electron density map (1.0 σ).

2.3.6 Confirmation of oligomeric state of Pgp3

Since crystals of Pgp3 contained one or three subunits in the asymmetric unit of wild-type and variants (H216A and H247A, both discussed later) depending on the space group, we analyzed the oligomeric states of the wild-type and variant Pgp3 in solution using SEC-MALS experiments. The molecular masses of the wild-type, H216A and H247A variants were 28.4, 28.9, and 30.0 kDa, respectively, close to the theoretical molecular mass of the Pgp3 monomer (29.4 kDa) (Figure 2.17). These results indicate that the likely functional state of Pgp3, as measured herein in solution, is monomeric. Interestingly, other structurally characterized cell-shape determining proteins from *H. pylori* may be either a monomer (Csd3) or dimer (Csd1 and Csd2) (An et al., 2016; Kirk et al., 2015).

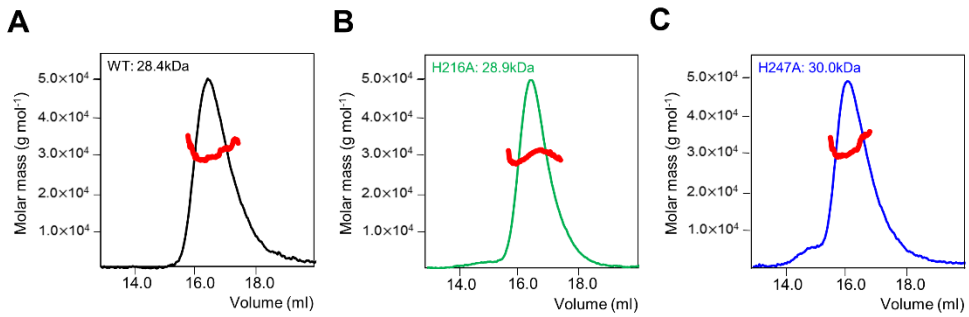


Figure 2.17 SEC-MALS experiment of wild-type, H216A, and H247A Pgp3 proteins.

SEC-MALS profiles for (A) wild-type Pgp3 (black), (B) H216A mutant (green), and (C) H247A mutant (blue). The thick red lines represent the measured molecular weight.

2.3.7 Structural insights into endolytic and exolytic reaction

To gain further structural insight into how Pgp3 shows both D,D-carboxypeptidase and D,D-endopeptidase activities, we examined how the mDAP^3-D-Ala^4-D-Ala^5 fragment from the pentapeptide and the mDAP^3-D-Ala^4-mDAP^3 fragment from the tetra-tri peptide is bound at the active site of Pgp3. The critical portions of both peptides with Pgp3 are shared motifs between the two structures. Both peptides contain the common motif, mDAP^3-D-Ala^4, and only the third amino acid is different, i.e., mDAP^3 vs Ala^5 (Figure 2.18). As described above (Figure 2.16), Tyr129, Arg195, and Tyr215 of Pgp3 contribute to peptide binding by forming several hydrogen-bonding networks with the first mDAP^3 moiety of either peptide.

However, the third amino acid of tetra-tri peptide and pentapeptide is different, i.e., mDAP^3 vs Ala^5 (Figure 2.18A), thus, we next examined that how the third amino acid of both tetra-tri peptide and pentapeptide is recognized by Pgp3. It is important to note that the third mDAP side chain of the tetra-tri peptide encompasses the "D-Ala" segment with the correct stereochemistry corresponding to D-Ala of pentapeptide (Figure 2.18B). Interestingly, two more interactions with backbone atoms of Ser244 and Val243 contribute for recognition of the third mDAP in tetra-tri peptide (Figure 2.16). Moreover, additional tripeptide-NAM-NAM moiety can be linked to N1 atom of the third mDAP of tetra-tri peptide because it is solvent-exposed in Pgp3 structure (shown in white-dotted circle in Figure 2.18B), suggesting that these structural features are crucial for duality of the endolytic and exolytic reactions of Pgp3.

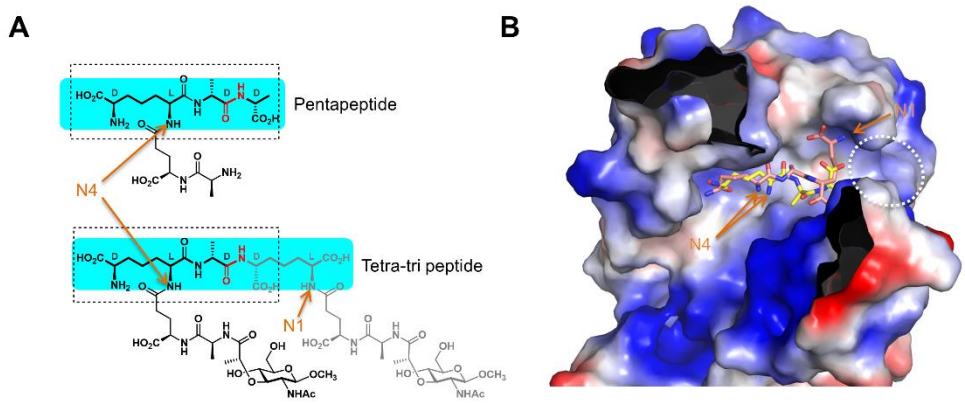


Figure 2.18 Structural comparison of binding modes of penta- and tetra-tri peptides to Pgp3.

(A) Alignment of the shared structural attributes for the pentapeptide and tetra-tri peptide, identified in the bound ligands. The scissile bonds are depicted in red. The shared structural component between the two substrates is in dotted box with the segments whose electron densities were observed highlighted in the shade boxes in cyan. (B) Surface representation of the electrostatic potential of Pgp3 in complex with pentapeptide (yellow) and tetra-tri peptide (pink). The positions of N1 and N4 atoms of each peptide are indicated by arrows. Solvent-exposed region next to the N1 atom of tetra-tri peptide is indicated by white dotted lines.

2.3.8 The catalytic mechanism

Our structural studies point to a catalytic mechanism for Pgp3 that is similar to that of other Zn^{2+}-dependent hydrolases (Figure 2.19) (Hudson et al., 1996; Lipscomb and Strater, 1996; Spencer et al., 2010). In the uncomplexed H247A variant, two water molecules (Wat-1 and Wat-2) occupy the active site (Figure 2.11C). Because Wat-1 and Wat-2 were essentially equidistant from Zn^{2+} (2.9 Å and 2.7 Å, respectively), we could not establish which should be the hydrolytic-water molecule. However, the crystal structure of the H247A variant in complex with the substrate clearly showed that Wat-2 was replaced by the incoming substrates, concluding that Wat-1 is the hydrolytic water that attacks the scissile carbonyl group of substrates. Substrate carbonyl oxygen displaces Wat-2, thereby interacting directly with Zn^{2+} and polarizing the D-Ala^4 carbonyl bond. This renders the amide bond susceptible to nucleophilic attack by Wat-1. The distance between Wat-1 and Zn^{2+} decreases from 2.7 Å in the uncomplexed H247A variant to 2.6 Å in the complexed one. The O3 atom of tartrate in wild-type Pgp3 (open form), which is the position of Wat-1 in H247A variant, is seen interacting with His216 and His247. Since these histidine residues are not involved in Zn^{2+} coordination, we hypothesized that either His216 or His247 might be able to abstract a proton from Wat-1 in promoting the hydrolytic-water molecule for attack at the susceptible substrate carbonyl carbon, en route to the stabilization of the transition state for the reaction within the enzyme oxyanion hole (Figure 2.19). Proton transfer from the histidine general base (His216 or His247) to the departing amide nitrogen facilitates cleavage of the peptide bond to generate a product complex in which the product carboxylate is bound analogously as in the crystallographically observed tartrate (Figure 2.19).

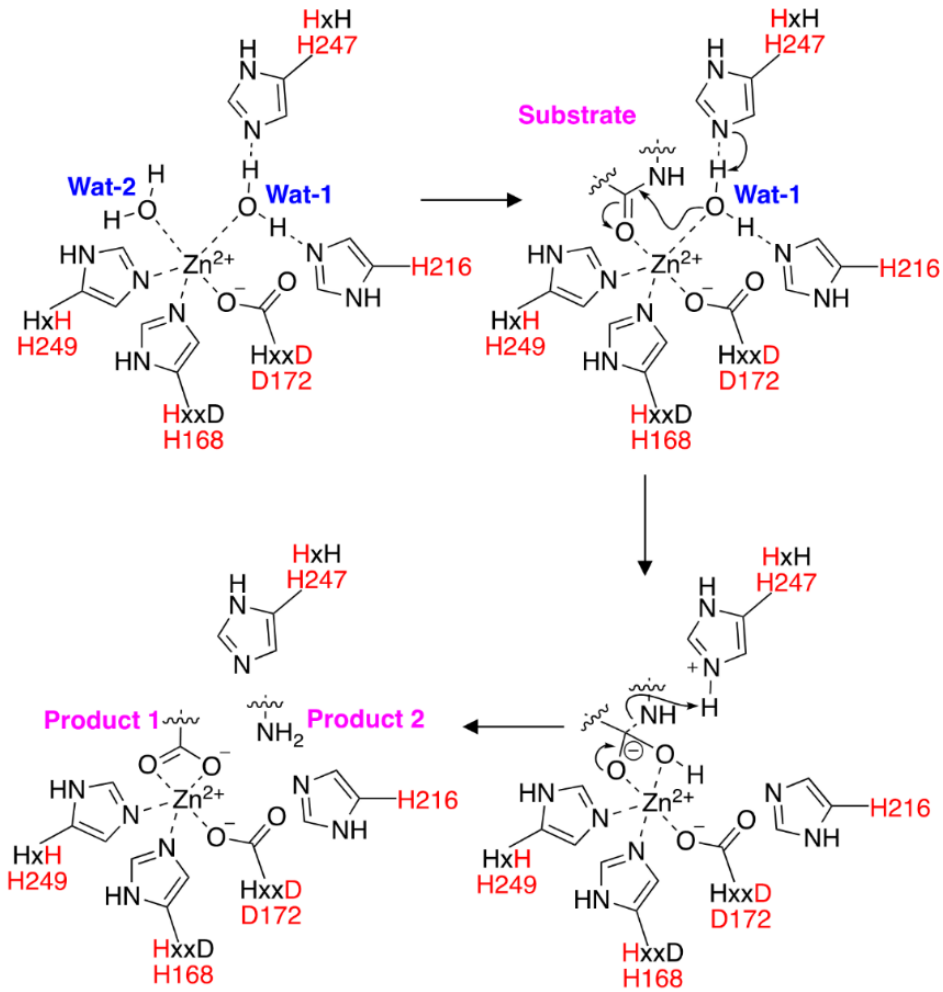


Figure 2.19 Proposed catalytic mechanism of Pgp3.

The proposed catalytic mechanism for turnover by Pgp3.

2.3.9 Computational modeling of Pgp3 with a peptidoglycan

As discussed above, Pgp3 adopts open and closed conformations due to the large conformational shift of the L1 loop (Figure 2.9). When we superimposed the structure of the H247A variant of Pgp3 in complex with the tetra-tri peptide with that of the wild-type complex with citrate (closed form), it was clear that the penta- or tetra-tri peptide cannot enter the active site in the closed conformation. The conclusion from our several structures is that catalysis in Pgp3 is regulated by the L1 loop.

The active site resides deep within the LytM domain. Importantly, the N1 and N4 atoms of the two mDAP of the tetra-tri peptide are responsible for linkage with the Glu^{2}-Ala^{1}-NAM moiety, which are solvent-exposed through the channels A and B, respectively (Figures 2.16 and 2.18). Due to the presence of the L1 loop, the access to the active site by the crosslinked peptidoglycan is prohibited from either the top or the bottom for steric reasons. Intuitively, the substrate could diffuse into the active site only sideways (channel A or B, Figure 2.16). The two intact polymeric glycan chains, which define the ends of the crosslinked peptide stem, also prohibit the sideways access of Pgp3. Pgp3 cannot access the intact cell-wall of *C. jejuni* without a prior processing of the peptide stem by another enzyme. For example, a different amidase/peptidase may cleave the peptide stem or the product of processing of the saccharide backbone by a lytic transglycosylase. Once this pre-requisite processing takes place, the resultant peptidoglycan, now tethered only to one polymeric glycan chain, can access the active site of Pgp3 from one of the channels. Interestingly, for the helical structure of bacteria, the amidases AmiA has a crucial role in cell-shape regulation by achieving the cleavage of N-acetylmuramoyl-L-alanyl bond (NAM-L-Ala^{1}) and regulating the degree of peptidoglycan cross-linking for the curved morphology (Chaput et al., 2016; Klockner et al., 2014). The AmiA reaction might serve as the pre-requisite reaction.

To determine which glycan end needs to be processed prior to the reaction of Pgp3, we characterized the approach of the substrate from either channel. We note that a narrow groove defines the surface of Pgp3 near channel A, which could ensconce the polysaccharide backbone from the donor strand in the course of the active-site penetration by the peptide stem (Figure 2.20A). We constructed computationally a full-length peptide stem (NAG-NAM(NAG)-L-Ala^1-γ-D-Glu^2-mDAP^3-D-Ala^4-mDAP^3-γ-D-Glu^2-Ala^1) within the active site based on the X-ray coordinates, which was submitted to molecular-dynamics simulations. The substrate formed a stable complex with Pgp3 during 250 ns of dynamics simulation, as discerned by its overall RMSD of 0.92 ± 0.23 Å (Figure 2.20B). His168, Asp172, His249 and D-Ala^4 carbonyl oxygen maintained a tetrahedral geometry with the Zn^{2+} ion during the MD simulation. The scissile amide bond transformed from the trans to cis configuration rapidly at the beginning of the simulation, revealing that catalysis favored the cis configuration of amide bond. Wat-1 also formed strong hydrogen bonds with His216 and His247. During MD, Wat-1 was properly aligned to transfer a proton to either His216 or His247 (average distance from Wat-1 O atom, 2.76 ± 0.04 Å and 2.71 ± 0.04 Å, respectively) prior to nucleophilic attack on the polarized carbonyl carbon of the substrate (average distance from Wat-1 O atom of 3.20 ± 0.04 Å). Tyr129, Arg195 and Tyr215 also consistently maintained hydrogen bonds and salt-bridge interactions with carboxyl group of mDAP^3 during MD simulation (occupancies of 15%, 35% and 36%, respectively). The carboxylate of the other mDAP^3 and of D-Ala^4 were seen interacting with the highly flexible Arg242 and Ser166 through salt-bridges and hydrogen bonds, respectively. One γ-D-Glu^2 could be seen interacting with Ser152, Lys156 and Ser244, while the other γ-D-Glu^2 formed intramolecular salt-bridge interactions with mDAP^3 (Figure 2.20C). The carbonyl group of L-Ala^1 formed a hydrogen bond with Arg158, which is present on the β8 strand (part of β7↑-β8↓-β9↑ near L1 linker

loop region). In the pentapeptide-bound complex, Arg158 might form salt-bridge interactions with D-Ala^5. Upon substrate binding, Arg158 extended into the active site and pulled L1 loop towards the active site. This indicates that the active site tends to close off after substrate binding. That is to say that the apo enzyme would sample both a closed and open conformations. For catalysis to proceed, the active site needs to open for the substrate to bind, but on binding, the loop closes again to enable catalysis, which then would release the two products from both channels A and B. This gives rise to the closed apo enzyme (seen in our structure), which is ready now to repeat the process.

We also analyzed the collective motions of Pgp3 with the help of temperature factors and elastic-network models. The L1 loop and loops between β12↓-β13↑ and β17↑-β18↓ play a role in the conformational dynamics. We also fine-tuned the conformational dynamics upon substrate binding, which we discerned in equilibrium dynamics, using guided sampling to expedite the conformational transition. We used temperature-Replica Exchange MD (tREMD) and Targeted MD (TMD). While tREMD utilizes temperature to cross the high-energy barrier between conformations, TMD uses RMSD between two pre-defined end states to drive the conformational change. Whereas tREMD sampled some of the conformational states in transition from one state to the other, TMD was more conclusive (smallest RMSD between the linker L1 in the closed conformation, and any MD snapshot in TMD and tREMD were 3.20 Å and 2.28 Å, respectively). Arg158 played a key role in pulling the β-sheet near L1 loop towards the substrate, which led to relaxing of the taut L1 loop, which eventually reorganized into a single-turn helix. A plot depicting RMSD between open and closed states during REMD and TMD is shown in Figure 2.21.

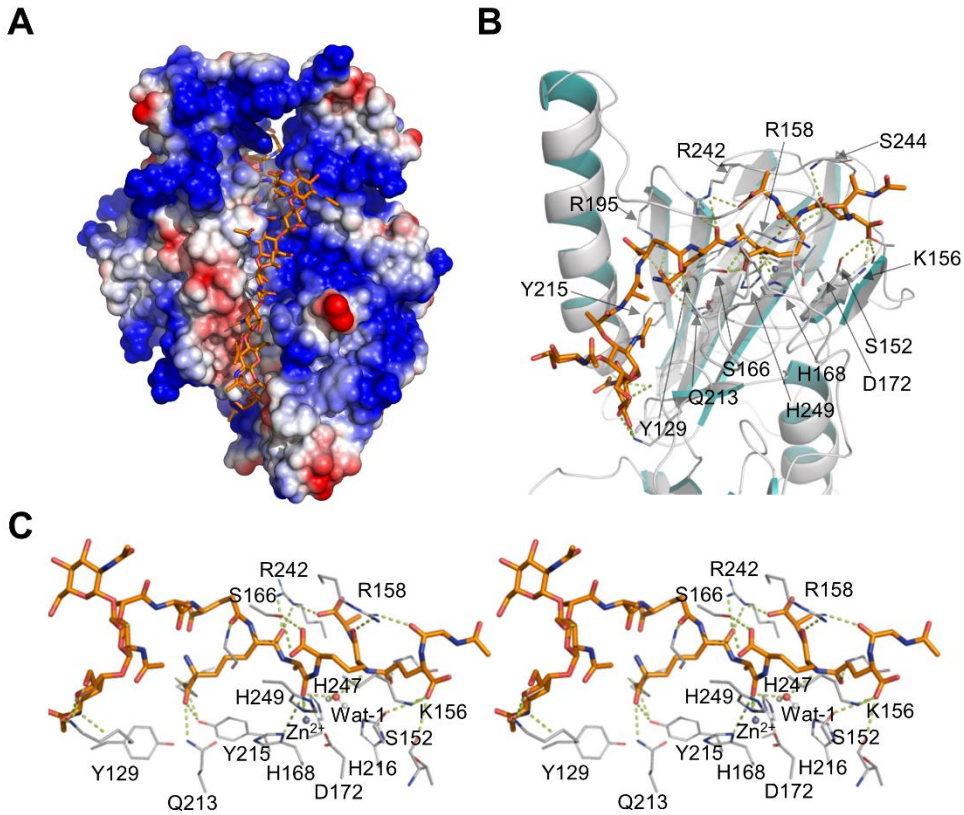


Figure 2.20 Predicted binding mode of peptidoglycan.

(A) A polymeric strand of peptidoglycan is shown ensconced in a groove leading to the active site. (B) Binding mode and interactions of the modeled peptidoglycan substrate in a snapshot of the 250-ns MD trajectory. Tyr129, Arg195, Gln213 and Tyr215 form significant interactions with the carboxyl group of mDAP[3]; Ser152 and Lys156 interact with carboxyl group of γ-D-Glu[2]; Arg158 interacts with γ-D-Glu[2]; Ser166 and Arg242 interact with carboxyl group of *m*DAP[3] and Ser244 interacts with γ-D-Glu[2]. (C) Stereo diagrams of the active-site bound peptidoglycan.

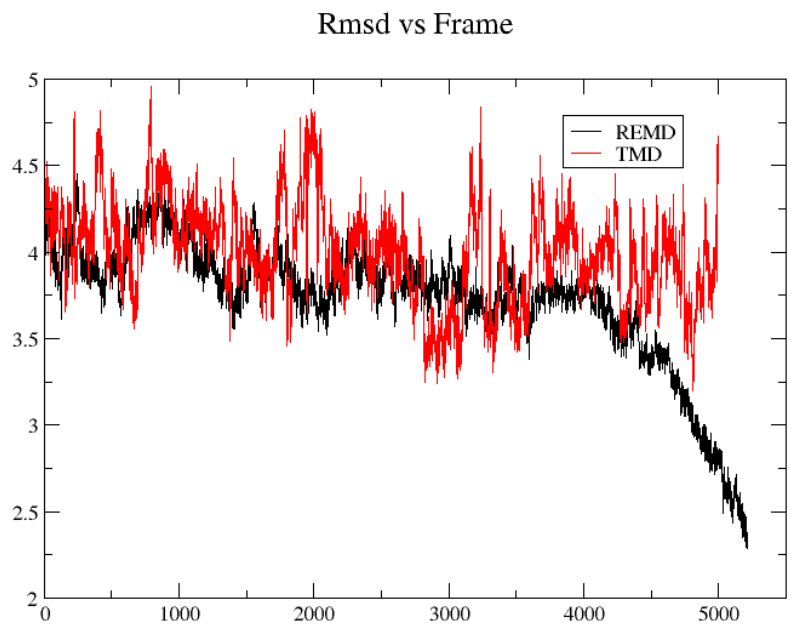


Figure 2.21 Conformational switch between open and closed states.

RMSD between open and closed states (regions: Glu99-Phe109, Thr159-Val164) in two MD techniques. TMD is more successful in obtaining the closed state from the open state as shown by their closing RMSD.

2.4 Discussions

We have described herein the reactions and the structure of Pgp3, a cell-shape-determining protein of *C. jejuni*. This enzyme turns over the cell-wall peptidoglycan by both the D,D-endopeptidase and D,D-carboxypeptidase activities, as documented by mass spectrometric characterization of the requisite reactions with synthetic cell-wall fragments. The high-resolution structures and molecular dynamics reveal an elaborate process of conformational change that samples the enzyme between active site closed and open states. Furthermore, based on the topographical construction of the active site, our work reveals that the peptidoglycan substrate for this enzyme can be tethered to the polymeric cell-wall saccharide backbone only on one strand (and not two). These are the structural and mechanistic insights into an important enzyme that regulates the shape and virulence of *C. jejuni*.

Our study can be utilized in designing an inhibitor for gastrointestinal pathogens such as *C. jejuni* and *H. pylori* by averting their helical cell-shape transformation. The inhibition of helical cell morphology might have a high impact on the invasion, motility, and infection ability of pathogens. From our structural studies, we observed that the citrate or tartrate molecule is tightly bound to the active site of Pgp3. In addition, we found that there are the common residues for the recognition of pentapeptide or tetra-tri peptide in the active site of Pgp3. These structural insights are feasibly useful in designing inhibitors interfering with gut colonization.

2.5 References

An, D.R., Im, H.N., Jang, J.Y., Kim, H.S., Kim, J., Yoon, H.J., Hesek, D., Lee, M., Mobashery, S., Kim, S.J., *et al.* (2016). Structural basis of the heterodimer formation between cell shape-determining proteins Csd1 and Csd2 from *Helicobacter pylori*. PLoS One ***11***, e0164243.

An, D.R., Kim, H.S., Kim, J., Im, H.N., Yoon, H.J., Yoon, J.Y., Jang, J.Y., Hesek, D., Lee, M., Mobashery, S., *et al.* (2015). Structure of Csd3 from *Helicobacter pylori*, a cell shape-determining metallopeptidase. Acta Crystallogr. D Biol. Crystallogr. ***71***, 675-686.

Blaser, M.J., and Engberg, J. (2008). Clinical Aspects of *Campylobacter jejuni* and *Campylobacter coli* Infections. In *Campylobacter* , Third Edition (American Society of Microbiology).

Brunger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature ***355***, 472-475.

Caccamo, P.D., and Brun, Y.V. (2018). The molecular basis of noncanonical bacterial morphology. Trends Microbiol. ***26***, 191-208.

Chaput, C., Ecobichon, C., Pouradier, N., Rousselle, J.C., Namane, A., and Boneca, I.G. (2016). Role of the N-acetylmuramoyl-l-alanyl amidase, AmiA, of *Helicobacter pylori* in peptidoglycan metabolism, daughter cell separation, and virulence. Microb. Drug Resist. ***22***, 477-486.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. ***66***, 12-21.

Costa, K., Bacher, G., Allmaier, G., Dominguez-Bello, M.G., Engstrand, L., Falk, P., de Pedro, M.A., and Garcia-del Portillo, F. (1999). The morphological transition of *Helicobacter pylori* cells from spiral to coccoid is preceded by a substantial modification of the cell wall. J.

Bacteriol. ***181***, 3710-3715.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. ***66***, 486-501.

Esson, D., Gupta, S., Bailey, D., Wigley, P., Wedley, A., Mather, A.E., Meric, G., Mastroeni, P., Sheppard, S.K., Thomson, N.R.*, et al.* (2017). Identification and initial characterisation of a protein involved in *Campylobacter jejuni* cell shape. Microb. Pathog. ***104***, 202-211.

Frirdich, E., Biboy, J., Adams, C., Lee, J., Ellermeier, J., Gielda, L.D., Dirita, V.J., Girardin, S.E., Vollmer, W., and Gaynor, E.C. (2012). Peptidoglycan-modifying enzyme Pgp1 is required for helical cell shape and pathogenicity traits in *Campylobacter jejuni*. PLoS Pathog. ***8***, e1002602.

Frirdich, E., Vermeulen, J., Biboy, J., Soares, F., Taveirne, M.E., Johnson, J.G., DiRita, V.J., Girardin, S.E., Vollmer, W., and Gaynor, E.C. (2014). Peptidoglycan LD-carboxypeptidase Pgp2 influences *Campylobacter jejuni* helical cell shape and pathogenic properties and provides the substrate for the DL-carboxypeptidase Pgp1. J. Biol. Chem. ***289***, 8007-8018.

Garrity, G.M., Bell, J.A., and Lilburn, T. (2005). Class V. *Epsilonproteobacteria* class. nov. In Bergey's Manual® of Systematic Bacteriology: Volume Two The Proteobacteria Part C The Alpha-, Beta-, Delta-, and Epsilonproteobacteria, D.J. Brenner, N.R. Krieg, and J.T. Staley, eds. (Boston, MA: Springer US), pp. 1145-1194.

Hermans, D., Van Deun, K., Martel, A., Van Immerseel, F., Messens, W., Heyndrickx, M., Haesebrouck, F., and Pasmans, F. (2011). Colonization factors of *Campylobacter jejuni* in the chicken gut. Vet. Res. ***42***, 82.

Hesek, D., Lee, M., Zhang, W., Noll, B.C., and Mobashery, S. (2009). Total

synthesis of N-acetylglucosamine-1,6-anhydro-N-acetylmuramylpentapeptide and evaluation of its turnover by AmpD from *Escherichia coli*. J. Am. Chem. Soc. ***131***, 5187-5193.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. ***38***, W545-549.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. ***233***, 123-138.

Hudson, K.R., Vernallis, A.B., and Heath, J.K. (1996). Characterization of the receptor binding sites of human leukemia inhibitory factor and creation of antagonists. J. Biol. Chem. ***271***, 11971-11978.

Huyton, T., Zhang, J.-G., Luo, C.S., Lou, M.-Z., Hilton, D.J., Nicola, N.A., and Garrett, T.P.J. (2007). An unusual cytokine:Ig-domain interaction revealed in the crystal structure of leukemia inhibitory factor (LIF) in complex with the LIF receptor. Proc. Natl. Acad. Sci. U S A ***104***, 12737.

Kirk, M.D., Pires, S.M., Black, R.E., Caipo, M., Crump, J.A., Devleesschauwer, B., Döpfer, D., Fazil, A., Fischer-Walker, C.L., Hald, T.*, et al.* (2015). World health organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. PLoS Med. ***12***, e1001921.

Klockner, A., Otten, C., Derouaux, A., Vollmer, W., Buhl, H., De Benedetti, S., Munch, D., Josten, M., Molleken, K., Sahl, H.G.*, et al.* (2014). AmiA is a penicillin target enzyme with dual activity in the intracellular pathogen *Chlamydia pneumoniae*. Nat. Commun. ***5***, 4201.

Lee, M., Hesek, D., Blazquez, B., Lastochkin, E., Boggess, B., Fisher, J.F., and Mobashery, S. (2015). Catalytic spectrum of the penicillin-binding protein 4 of *Pseudomonas aeruginosa*, a nexus for the induction of beta-lactam antibiotic resistance. J. Am. Chem. Soc. ***137***, 190-200.

Lee, M., Hesek, D., Dik, D.A., Fishovitz, J., Lastochkin, E., Boggess, B., Fisher,

J.F., and Mobashery, S. (2017) From genome to proteome to elucidation of reactions for all eleven known lytic transglycosylases from *Pseudomonas aeruginosa.* Angew.Chem. Int. Ed. ***56***, 2735-2739.

Lee, M., Hesek, D., Shah, I.M., Oliver, A.G., Dworkin, J., and Mobashery, S. (2010). Synthetic peptidoglycan motifs for germination of bacterial spores. Chembiochem ***11***, 2525-2529.

Lipscomb, W.N., and Strater, N. (1996). Recent advances in zinc enzymology. Chem. Rev. ***96***, 2375-2434.

Meroueh, S.O., Bencze, K.Z., Hesek, D., Lee, M., Fisher, J.F., Stemmler, T.L., and Mobashery, S. (2006). Three-dimensional structure of the bacterial cell wall peptidoglycan. Proc. Natl. Acad. Sci. U S A ***103***, 4404-4409.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. ***53***, 240-255.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. ***276***, 307-326.

Sauvage, E., Kerff, F., Terrak, M., Ayala, J.A., and Charlier, P. (2008). The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol. Rev. ***32***, 234-258.

Scheffers, D.J., and Pinho, M.G. (2005). Bacterial cell wall synthesis: new insights from localization studies. Microbiol. Mol. Biol. Rev. ***69***, 585-607.

Spencer, J., Murphy, L.M., Conners, R., Sessions, R.B., and Gamblin, S.J. (2010). Crystal structure of the LasA virulence factor from *Pseudomonas aeruginosa*: substrate specificity and mechanism of M23 metallopeptidases. J. Mol. Biol. ***396***, 908-923.

Stahl, M., Frirdich, E., Vermeulen, J., Badayeva, Y., Li, X., Vallance, B.A., and Gaynor, E.C. (2016). The helical shape of *Campylobacter jejuni*

promotes in vivo pathogenesis by aiding transit through intestinal mucus and colonization of crypts. Infect. Immun. ***84***, 3399-3407.

Stahl, M., Ries, J., Vermeulen, J., Yang, H., Sham, H.P., Crowley, S.M., Badayeva, Y., Turvey, S.E., Gaynor, E.C., Li, X.*, et al.* (2014). A novel mouse model of *Campylobacter jejuni* gastroenteritis reveals key pro-inflammatory and tissue protective roles for Toll-like receptor signaling during infection. PLoS Pathog. ***10***, e1004264.

Suvorov, M., Lee, M., Hesek, D., Boggess, B., and Mobashery, S. (2008). Lytic Transglycosylase MltB of *Escherichia coli* and its role in recycling of peptidoglycan strands of bacterial cell wall. J. Am. Chem. Soc. ***130***, 11878-11879.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. ***66***, 22-25.

Vollmer, W., and Bertsche, U. (2008). Murein (peptidoglycan) structure, architecture and biosynthesis in *Escherichia coli*. Biochim. Biophys. Acta ***1778***, 1714-1734.

Vollmer, W., Blanot, D., and de Pedro, M.A. (2008). Peptidoglycan structure and architecture. FEMS Microbiol. Rev. ***32***, 149-167.

Wassenaar, T.M., van der Zeijst, B.A., Ayling, R., and Newell, D.G. (1993). Colonization of chicks by motility mutants of *Campylobacter jejuni* demonstrates the importance of flagellin A expression. J. Gen. Microbiol. ***139*** *Pt 6*, 1171-1175.

Wyckoff, T.J., Taylor, J.A., and Salama, N.R. (2012). Beyond growth: novel functions for bacterial cell wall hydrolases. Trends Microbiol. ***20***, 540-547.

Yang, D.C., Blair, K.M., and Salama, N.R. (2016). Staying in shape: the impact of cell shape on bacterial survival in diverse environments. Microbiol. Mol. Biol. Rev. ***80***, 187.

국문 초록

β-arrestin2와 CXCR7의 C-말단 인산화 펩타이드 복합체 구조 규명을 통한 β-arrestin2의 활성 기작 연구 및 장내 병원성 세균인 *Campylobacter jejuni*의 나선형 세포 모양을 결정하는 Pgp3 단백질의 구조 및 기능 연구

민경진

자연과학대학 화학부

서울대학교 대학원

본 논문은 리간드 인식이 단백질 구조 및 기능의 변화에 어떻게 영향을 미치는지에 대한 분자 수준의 이해를 얻는 데 중점을 둔다. 본 논문은 두 개의 장으로 나뉜다. 제1장에서는 G-단백질 결합 수용체 (GPCR) 인식에 대한 분자 수준의 이해를 얻기 위해 중요한 세포 신호 조절 단백질인 β-arrestin (βarr)을 설명한다. 제2장에서는 *C. jejuni*의 세포 모양 결정 (Csd) 단백질 중 하나인 펩티도글리칸 분해효소 3 (Pgp3)에 대해 논의한다. Pgp3 단백질의 펩티도글리칸 인식에 대한 분자 수준의 연구는 병원체의 숙주 감염 메커니즘에 대한 통찰력을 갖게 할 것이다.

제1장에서는 βarr2의 구조적 및 생물 물리학적 연구에 대해 다룬다. βarr은 GPCR의 신호 전달 및 이동을 조절하는데 중요하며, βarr은 βarr1과 βarr2 이렇게 두 가지 동위체가 존재한다. GPCR은 세포막 수용체 중에서 가장 큰 패밀리를 가진 수용체이며, 약물 표적에 주요 단백질이다. G-단백질 매개 GPCR 신호 전달을 끄기 위해서, GPCR 인산화 효소는 GPCR의 C-말단 꼬리 및/또는 세포 내 존재하는 루프를 인산화 시켜, βarr와의 결합을

초래한다. 현재까지, GPCR의 C-말단 꼬리로부터 유래된 인산화 펩타이드와 활성형 (activated) βarr1의 고해상도 복합체 구조가 밝혀졌지만, βarr2와의 복합체 구조는 밝혀져 있지 않았다. βarr2 단백질에 의한 GPCR의 인산화 된 카복실-말단의 인식을 분자적으로 이해하기 위해, class A GPCR인 CXCR7 수용체의 C-말단에서 유래된 인산화 펩타이드 (C7pp)와 βarr2 단백질의 복합체 결정 구조를 규명하였다. C7pp와 결합된 βarr2의 구조는 기존에 밝혀진 βarr1 활성형 구조와 주요한 차이점을 보인다. 주요 차이점 중 하나는 C7pp와 결합된 βarr2의 구조가 비교적 작은 도메인 (N-도메인과 C-도메인) 간 회전을 나타낸다는 것이다. 또한, 본 연구에서 규명한 βarr2의 CXCR7 인산화 펩타이드 복합체 구조가 활성형임을 확인하기 위하여 수소/중수소 교환 질량 분석법 (HDX-MS) 및 합성 항체 기반 입체 구조 센서를 사용하였다.

제2장에서는 *C. jejuni*의 Pgp3 단백질에 대한 구조적 및 기능적 연구가 논의된다. *C. jejuni*의 나선형 세포 모양은 사람의 장 내 감염을 위해 박테리아 군집형성에 중요하며, 나선형 세포 모양은 펩티도글리칸의 특정 유형의 가교 결합으로 인한 것으로 여겨진다. Csd 단백질에 의한 *C. jejuni*의 펩티도글리칸 리모델링을 통해, *C. jejuni*는 박테리아 군집화 및 발병 기전에 적합하도록 세포의 모양을 유지하게 된다. Csd 단백질이 펩티도글리칸을 인식하는 방법을 이해하기 위해, 서로 다른 2가지의 펩티도글리칸 유도체 기질이 결합된 서로 다른 복합체 구조 2개를 포함하여 총 8개의 Pgp3 단백질의 X-선 결정 구조를 규명하였다. 또한, Pgp3 단백질은 펩티도글리칸 세포벽에 대해 D,D-endopeptidase와 D,D-carboxypeptidase 활성이 있음을 밝혔다. Pgp3의 촉매 작용은 펩티도글리칸에 결합할 때 큰 구조적 변화를 동반하는 것으로 나타났는데, 특히 Pgp3의 루프가 기질이 활성 부위에 접근하는데 조절함을 밝혔다. 또한, 한쪽의 펩티도글리칸의 당류 골격으로부터의 가교 된 펩티드 줄기의 가수 분해는 Pgp3에 의한

펩티도글리칸의 인식 및 촉매 작용을 위한 전제 조건임을 유추할 수 있었다. 이러한 분석은 사람의 장내 병원체인 *C. jejuni*의 병원성을 정의하는 핵심이 세포 모양 형태라는 점에서 비전형적 특성을 나타낸다는 것을 말한다.

주요어: β-arrestin2, CXCR7, 세포 모양 결정 단백질, 펩티도글리칸, 병원균, X-선 결정학

학번: 2015-22612